

**Ques:el/Orbit**

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2000, 19:38:00 ; Search time 1395.77 Seconds  
(without alignments)  
3745.598 Million cell updates/sec

Title: US-08-774-104A-1

Perfect score: 1197  
Sequence: 1 ATGATGACATCATCGAAGC.....ACTACATCATGTCGATGA 1197

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_p11:\*  
8: gb\_p12:\*  
9: gb\_p11:\*  
10: gb\_p12:\*  
11: gb\_p13:\*  
12: gb\_ro:\*  
13: gb\_sy:\*  
14: gb\_un:\*  
15: em\_fun:\*  
16: em\_hum1:\*  
17: em\_hum2:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: gb\_htg1:\*  
31: gb\_htg2:\*  
32: gb\_in1:\*  
33: gb\_in2:\*  
34: em\_ba1:\*  
35: em\_ba2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: gb\_p14:\*  
39: gb\_htg3:\*  
40: gb\_htg4:\*  
41: gb\_htg5:\*  
42: gb\_htg6:\*  
43: gb\_htg7:\*

44: em\_htg1:\*  
45: em\_htg2:\*  
46: em\_htg3:\*  
47: em\_hum5:\*  
48: gb\_p13:\*  
49: gb\_p15:\*  
50: gb\_htg8:\*  
51: gb\_htg9:\*  
52: gb\_htg10:\*  
53: gb\_htg11:\*  
54: gb\_htg12:\*  
55: gb\_htg13:\*  
56: gb\_htg14:\*  
57: gb\_in3:\*  
58: gb\_htg15:\*  
59: gb\_htg16:\*  
60: gb\_htg17:\*  
61: em\_htg4:\*  
62: em\_htg5:\*  
63: em\_htg6:\*  
64: em\_htg7:\*  
65: em\_hum6:\*  
66: gb\_htg18:\*  
67: gb\_htg19:\*  
68: gb\_htg20:\*  
69: gb\_htg21:\*  
70: gb\_htg22:\*  
71: gb\_htg23:\*  
72: gb\_v11:\*  
73: gb\_v12:\*  
74: gb\_ba3:\*  
75: em\_htg8:\*  
76: em\_htg9:\*  
77: em\_htg10:\*  
78: em\_htg11:\*  
79: em\_htg12:\*  
80: em\_htg13:\*  
81: em\_htg14:\*  
82: em\_htg15:\*  
83: em\_htg16:\*  
84: em\_htg17:\*  
85: em\_htg18:\*  
86: em\_htg19:\*  
87: em\_htg20:\*  
88: em\_htg21:\*  
89: em\_htg22:\*  
90: em\_htg23:\*  
91: gb\_p16:\*  
92: gb\_p17:\*  
93: gb\_sts1:\*  
94: gb\_sts2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1197	100.0	1197	11	AF069740	AF069740 Homo sapi
2	108.6	9.1	5932	2	BACMENAOUI	M74521 Bacillus su
3	108.6	9.1	5946	2	BACMENAOOP	M74538 Bacillus su
c 4	108.6	9.1	215640	2	BSUB0016	299119 Bacillus su
5	108.6	9.1	220060	1	AF008220	AF008220 Bacillus
6	103.4	8.6	4662	2	PAX82644	X82644 P.aeruginos
7	89.8	7.5	143051	2	D90911	D90911 Synechocyst
8	76	6.3	1659	2	ECOE011	M24142 E. coli iso
9	76	6.3	12003	1	AE000165	AE000165 Escherich
10	76	6.3	136742	2	ECU82598	U82598 Escherichla
c 11	74	6.2	14027	1	AE004272	AE004272 Vibrio ch
12	72.8	6.1	1302	2	ECOE011	M36700 E. coli ito

13	60.8	5.1	1560	2	ECU54790	U54790 Escherichia
14	59.8	5.0	1973	7	AF078080	AF078080 Arabidops
15	57	4.8	2078	8	CRO6065	AJ006065 Catharant
16	51	4.3	1358	2	ECMENFGN	U50849 E.coli menf
17	51	4.3	10409	1	AEO000316	AEO000316 Escherich
18	51	4.3	13348	8	D90858	D90858 E.coli geno
19	51	4.3	15676	2	D90857	D90857 E.coli geno
20	49.8	4.2	5166	2	BSU26644	U26444 Bacillus su
21	49.8	4.2	217420	2	BSU00017	U29120 Bacillus su
22	48	4.0	44880	2	BACRFO01A	M34053 B.subtilis
23	48	4.0	180136	1	BAC180K	D26185 B. subtilis
24	48	4.0	213080	2	BSUB0001	U29104 Bacillus su
25	47.8	4.0	110000	50	AC020850_1	Continuacion (2 o
26	46.6	3.9	22070	2	MTCY7511	U295120 Mycobacteri
27	46	3.8	164921	7	AF0742186	AF074221 Mus musc
28	45	3.8	332375	60	AF0074221	AF0074221 Mus musc
29	44.8	3.7	5057	2	PFBSCEAB	Y03356 Pseudomonas
30	44.4	3.7	8656	2	LACTROP	M87483 L. lactis t
31	44	3.7	12048	1	AEO004162	AEO004162 Vibrio cho
32	44	3.7	122000	74	VCU52150	U52150 Vibrio cho
33	43.6	3.6	2591	2	L81122	L81122 Buchnera ap
34	43.4	3.6	2882	2	SSOSRRPFG	M96048 Sulfolobus
35	43.4	3.6	6380	2	SSRRPFG	Z50014 S.solfatar
36	42.4	3.5	351	74	YEN286281	AJ286281 Yersinia
37	41.4	3.5	293350	2	CNSPAX02	AJ274049 Yersinia
38	41	3.4	351	74	YEN270409	AJ286282 Yersinia
39	41	3.4	351	74	YEN288282	AJ286284 Yersinia
40	41	3.4	351	74	YEN288284	AJ288285 Yersinia
41	41	3.4	351	74	YEN288285	U01378 Salmonella
42	41	3.4	1563	74	STRPPE	U01812 S.typhimur
43	41	3.4	5621	32	AF030155	AF030155 Drosophila
44	40.6	3.4	5621	32	AF030155	AF030152 Drosophila
45	40.6	3.4	13548	41	AC015023	AC015023 Drosophila

## ALIGNMENTS

FEATURES	source
LOCUS 1	AF069740
DEFINITION	AF069740 Homo sapiens neutral sphingomyelinase mRNA, complete cds.
ACCESSION	AF069740
VERSION	AF069740.1 GI:6606079
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 1197)
TITLE	Chatterjee,S., Han,H., Rollins,S. and Cleveland,T.
JOURNAL	Molecular cloning, characterization, and expression of a novel human neutral sphingomyelinase
MEDLINE	J. Biol. Chem. 274 (52), 37407-37412 (1999)
REFERENCE	20669736
AUTHORS	2 (bases 1 to 1197)
TITLE	Chatterjee,S., Han,H., Rollins,S. and Cleveland,T.
JOURNAL	Direct Submissions
Submitted (03-JUN-1998)	Pediatrics, Johns Hopkins University School of Medicine, 600 N. Wolfe Street, CMSC 6-124, Baltimore, MD 21287-3654, USA
location/Qualifiers	
1..1197	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/tissue_type="kidney"	
1..1197	
/codon_start=1	
/product="neutral sphingomyelinase"	
/protein_id="A019052.1"	
/db_xref="GI:6606080"	
/translation="MMYHETRLAOSDLQQLYAALTEFFGAYFAIPADDTLRFGIGATAATTAQALGAVVGAQSPDEQETPSELNAGFEWFEVVEVNTIADKTKTFSSDPIV	

Query Match	Best Local Similarity	100.0%; Score 1197; DB 11; Length 1197;
Matches 1197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY 1 ATGATGACATATCAAGAAACGGCGCGCTTGGCTCAAAAGCAGCTTACAGCAACTCTATGCG 60	100.0%; Score 1197; DB 11; Length 1197;	SDFTTHLAQFVPRKQPNVTVTSHTYDEVDMLERFENLIDTLAIDQTLAKVYFGKQITLQ
DB 1 ATGATGACATATCAAGAAACGGCGCGCTTGGCTCAAAAGCAGCTTACAGCAACTCTATGCG 60		LSDTLAQIITIRALAEQANTYHYVVKHVDLFIATPERLVAAASGGGATPAVAATSR
QY 61 GCACCTTGAACAACACTGAATTTGGCGCTTACTTTGGGACACCCGCTGATGATCTTTACGT 120		RGSDGADDIALGEGALLLSOKNRHIEHQVBSLITTRDLVTSLSKVPMPSPILNLDVO
DB 61 GCACCTTGAACAACACTGAATTTGGCGCTTACTTTGGGACACCCGCTGATGATCTTTACGT 120		HLVYPIGTGIDIAHLSTVAIDRLPTPALGSGVREALYIAIATHEKIPRLPFGPIGY
QY 121 TTTGGCATTTGGCGCAATGCTACGGCAAAAACGGCTCGAGCATTTACAAAGGTGGCGTTGTT 180		FTANSGEFPVGIKSMVYVQTORATLTFAGAGIYADSDAQOEVEETGLKREPKRQLLK
DB 121 TTTGGCATTTGGCGCAATGCTACGGCAAAAACGGCTCGAGCATTTACAAAGGTGGCGTTGTT 180		DYNHVE"
QY 181 TTTGGTGGCGAGTCATTTGATGATCAACAGATACCCGCGAGTCATGTAATTGATGGCGGTTTT 240		
DB 181 TTTGGTGGCGAGTCATTTGATGATCAACAGATACCCGCGAGTCATGTAATTGATGGCGGTTTT 240		
QY 241 TGGTTTGTCCCGAAGTATGCTGACCATTCGCGCGAGATAAAAATCAGTTCGATCAGAT 300		
DB 241 TGGTTTGTCCCGAAGTATGCTGACCATTCGCGCGAGATAAAAATCAGTTCGATCAGAT 300		
QY 301 ACCGATATGATTTTATAGACGGTGGCGGAGTTCGTCGCAAAAACGCCAAATACGGGTG 360		
DB 301 ACCGATATGATTTTATAGACGGTGGCGGAGTTCGTCGCAAAAACGCCAAATACGGGTG 360		
QY 361 ACCACTACTCATGTGACAGATGAAAGTGATTCGAAACGACAGAGAAATTTGATTGAT 420		
DB 361 ACCACTACTCATGTGACAGATGAAAGTGATTCGAAACGACAGAGAAATTTGATTGAT 420		
QY 421 ACCTTAGCCATCGATCAAACTTAGCCAAAGTCGTTTTGGTGGGCAACAGACCCCTGAG 480		
DB 421 ACCTTAGCCATCGATCAAACTTAGCCAAAGTCGTTTTGGTGGGCAACAGACCCCTGAG 480		
QY 481 TTATCCGACAGTTACAGACTGGCAAAATTTGTCGCTGATACGTCAGGACAGGCAATACG 540		
DB 481 TTATCCGACAGTTACAGACTGGCAAAATTTGTCGCTGATACGTCAGGACAGGCAATACG 540		
QY 541 TATCATGTGATTTTAAACGACATGATGAATTTGTTTATTTACGAAACACCGGAACGGTTA 600		
DB 541 TATCATGTGATTTTAAACGACATGATGAATTTGTTTATTTACGAAACACCGGAACGGTTA 600		
QY 601 GTGGCTATGTAGGTTGTCAGATCGCTACGGCGGCGGTCGCTGGGACAAAGCCGGCGGG 660		
DB 601 GTGGCTATGTAGGTTGTCAGATCGCTACGGCGGCGGTCGCTGGGACAAAGCCGGCGGG 660		
QY 661 ACGGATGGCGCTGACAGATATCGCGTTAGGCGCAAGCGTTGTTAGCCAGTCAGAAAAACCGC 720		
DB 661 ACGGATGGCGCTGACAGATATCGCGTTAGGCGCAAGCGTTGTTAGCCAGTCAGAAAAACCGC 720		
QY 721 ATTGAACATCAATATGTCGTGGCAAGTATCAAGACAGCTTGAAGAAGTACGACAGCGTG 780		
DB 721 ATTGAACATCAATATGTCGTGGCAAGTATCAAGACAGCTTGAAGAAGTACGACAGCGTG 780		
QY 781 CTTAAAGTGCCGCGCATGCCCAAGTTTACTCAAAAATAGCAAGTTTCAGCATTTGTACACA 840		
DB 781 CTTAAAGTGCCGCGCATGCCCAAGTTTACTCAAAAATAGCAAGTTTCAGCATTTGTACACA 840		
QY 841 CCAATTTACAGGGGACATTTGGCGCAATTTAAAGTGTACCCGAGATTGTGACCCGCTTGAT 900		
DB 841 CCAATTTACAGGGGACATTTGGCGCAATTTAAAGTGTACCCGAGATTGTGACCCGCTTGAT 900		

QY	901	CCACACACAGACACTGGGAGGCGTCCACAGTGAAGGCGCCGTATATACATGGCAGCCAT	960
Db	901	CCAAACACAGACACTGGGAGGCGTCCACAGTGAAGGCGCCGTATATACATGGCAGCCAT	960
QY	961	GAGAAGACACCTCGGCGCTTGTTCACAGCTCATTTGGCTATTTTACCGAGATATAGT	1020
Db	961	GAGAAGACACCTCGGCGCTTGTTCACAGCTCATTTGGCTATTTTACCGAGATATAGT	1020
QY	1021	GGGGAATTTGTGGTGGCATCCGTTCCATGTATGTAATCAAAACGACGACGACCAACT	1080
Db	1021	GGGGAATTTGTGGTGGCATCCGTTCCATGTATGTAATCAAAACGACGACGACCAACT	1080
QY	1081	TTATTTGCTGGGCGCGGATTTGGCTACACCCGATCGGCGCAACAAGATATGAAGAACT	1140
Db	1081	TTATTTGCTGGGCGCGGATTTGGCTACACCCGATCGGCGCAACAAGATATGAAGAACT	1140
QY	1141	GGGTTGAATTTGAACCCATGGCGCAATTTGTTAAAGACTACATCATGTGCAATGA	1197
Db	1141	GGGTTGAATTTGAACCCATGGCGCAATTTGTTAAAGACTACATCATGTGCAATGA	1197
RESULT	2		
BACMENAQUI			
LOCUS	BACMENAQUI	5932 bp	DNA
DEFINITION	Bacillus subtilis menaquinone operon, complete cds.	BCP	12-OCT-1994
ACCESSION	M74521		
VERSION	M74521.1	GI:557486	
KEYWORDS	menaquinone.		
SOURCE	Bacillus subtilis		
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus.		
REFERENCE	1 (bases 1 to 4000; 4001 to 5932)		
AUTHORS	Driscoll,J.R. and Taber,H.W.		
TITLE	Sequence organization and regulation of the Bacillus subtilis member operon		
JOURNAL	J. Bacteriol.	174 (15),	5063-5071 (1992)
MEDLINE	92332443		
FEATURES			
Source	Location/Qualifiers		
	1..5932		
	/organism="Bacillus subtilis"		
	/db_xref="taxon:1423"		
	/germline		
	13..27		
	/note="putative"		
	/function="Possible regulatory site for pH regulation of promoter"		
	13..18		
	35..40		
	47..>579		
	/function="Message containing ORF1 to possibly ORF5"		
	/evidence=experimental		
	128..133		
	/note="putative"		
	142..579		
	/gene="ORF1"		
	142..579		
	/gene="ORF1"		
	142..579		
	/gene="ORF1"		
	/standard_name="ORF1"		
	/note="putative"		
	/codon_start=1		
	/transl_table=11		
	/function="unknown"		
	/product="unknown"		
	/protein_id="AA50396.1"		
	/db_xref="GI:557487"		
	/translation="MTYTVORTFRKKVLHLAKAEVNHAVLISYHQISLDPLSEFF		
	NYGKATKGNFEWSDPESELTIVGLKEAVFQIKNRSERYREVFQWMEFKTAAH		
	ITYEKKIOMDCYSEDFLLTLAKKEVINGTISRKISLCLRLC"		
	582..1547		
	/gene="ORF2"		
	582..1547		
CDS			

```

/gene="ORF2"
/standard_name="ORF2"
/note="putative"
/codon_start=1
/transl_table=11
/function="unknown"
/product="unknown"
/protein_id="AA50397.1"
/db_xref="GI:357488"
/translacion="MTAGPEPLTVNRWVGSGEDAENVLEGLKAFPAEFNVPDEKQEDQ
AVIAAEELEDDMDWLKAIEPTASQIKEQYDKVYLARELLTFDGPLOIEPVLTGLD
DOGTVEALIDQEGKTEFWGASPERLIRKQDGTWVSSGLASIRGVNEEDRRIGLEL
IDDLLENLHEHIVGMHNAFVSSCEVKEKQDGVLTAKTSVQHLPPVIGOLRESAS
IPDLERKHPPLPALGDSFOEKAVDYIRKEIPMSGVMAATIGMIDSDNNEFVAIRS
GLIEGSTAALRESQGIVEDSEPISEYETQIKLPMTALUGERR"
1624. .2049
/gene="menc"
1624. .2049
/gene="mencF"
/standard_name="2succiny]-6hydroxy-2,
4cyclohexadiene-1carboxylate synthase"
/codon_start=1
/transl_table=11
/function="Convert isochorismate to SHCHC"
/evidence=experimental
/protein_id="AA50398.1"
/db_xref="GI:357489"
/translacion="MSGIKIYTLAVLCAHPDISVHVOIDERSAGFPALGLAKAKORP
VLLITGSTAANEFYPAVEAHYSRVPILVLTADRPHELRRVGAPLPQINQFLGFMFV
KEFTDSALPESPOLKRIYRTLASRAGEAKRRMALSM"
2072. .3187
/gene="mencP"
2072. .3187
/gene="mencF"
/standard_name="mencF"
/codon_start=1
/transl_table=11
/evidence=experimental
/protein_id="AA50399.1"
/db_xref="GI:357490"
/translacion="MPDISDEPFGMRMTGRHVSVKTGTOSVDRESLSDVAEMLEAEK
GMIVGGEHLSADKENITATLSKALQYPLIADPLSNLNGVHXDSYTDAYDSTLKDDE
LKRKRIPVAVLIRFGPMPSKPEVFLMKDDPTIQIIVIDEGGWRDPTQASAHMHCNA
SVFAEIAAGLTAATRSSEMLEKQFPVNGRREHLQIISSEDSVEGNLRILOHLVP
ENSSLFSVNSMPIRDVDFPEKODRPFRIYSNRGANGIDGVSSAMGVCGCTAPVLT
VYGDLSFVHNDLNGLLAAKKKIGLPIVLIVNNDGGGIFSPLPQASEKTHFEDLFGPTLG
IDFKRIAAALYCGTYSGPCASMDPEFTATARROTSPDSI"
3270. .3857
/gene="ORF5"
3270. .3857
/gene="ORF5"
/standard_name="ORF5"
/note="putative"
/codon_start=1
/transl_table=11
/function="unknown"
/product="unknown"
/protein_id="AAA50400.1"
/db_xref="GI:357491"
/translacion="WGTVNITVSDGVRVYAADEGPNASAEAVVCLIHGFTKSGOSWTFLD
EMLPDRSLIKIDCLQHGTEADPLNGKRYSTTRQVSDLAIEFDLKLHKVLIQYSMGG
RLATYFAMTYVSGYHCLICAKRRGSKHLGNGNESCGETSLILTFEYTSKRILRIG
RLSLFHLHSRGKLIFFGCVYBRAWCGTIRGLRTV"
4089. .4103
/note="putative"
/function="Sequence may be involved in pH regulation of
promoter."
/phenotype="Unknown"
4102. .4107
4125. .4130
4138. .-57722
/gene="menB"
/evidence=experimental

```

BASE COUNT	1594	a	1248	c	1552	g	1538	t	ORIGIN
gene	4138..	.5722							
	/gene="menB"								
RBS	4148..	.4152							
	/gene="menB"								
	/note="putative"								
CDS	4157..	.4942							
	/gene="menB"								
	/standard_name="Men B"								
	/codon_start=1								
	/transl_table=1								
	/function="Convert OSB-CoA to DNA in Menaquinone biosynthesis"								
	/evidence=experimental								
	/product="dihydroxyapthoic acid (DHNA) synthetase"								
	/protein_id="AAAS0401.1"								
	/db_xref="GI:557492"								
	/translation="MKIFDETYNGIAKITTINRPVHNAFTPKVTAEMIDAFADARDQ								
	NWGYVLGAGADKAFCSGQDKVRGSGYVDDQIPRLNVLQRLIRVIRPKPVAAIM								
	SGVYIGGHHVLIYCDLTIADNAPIFGOTGPKVSGFAGYSGYLARIVGKKRKRETIM								
	YLICQYNAQEAALDMGIVNTVYPLPOLPEERTIKMCEMEIEKSPALRPLKAFNADTDG								
	LAGIQQFAGDATTLYITTTDEAKEGRDSFKERKPFQFPFP"								
	4963..	.4993							
terminator	/gene="menB"								
	/evidence=experimental								
	4975..	.4982							
misc_feature	/gene="menB"								
	/standard_name="Golden Loop' sequence"								
	/function="possible stabilized of RNA secondary structure."								
	/evidence=experimental								
	/product="None"								
	4986..	.5013							
misc_signal	/gene="menB"								
	/note="putative"								
	/function="possible recognition sequence for pH regulation of promoter."								
	/phenotype="Unknown"								
	5002..	.5007							
-35_signal	/gene="menB"								
	5024..	.5029							
	/gene="menB"								
-10_signal	5038..	.>5932							
mRNA	/gene="menB"								
	/evidence=experimental								
	5038..	.5932							
gene	/gene="menB"								
	5051..	.5056							
RBS	/gene="menB"								
	/note="putative"								
	5060..	.5722							
CDS	/gene="menB"								
	/standard_name="MenB"								
	/codon_start=1								
	/transl_table=1								
	/function="Convert OSB to OSB-CoA in Menaquinone biosynthesis."								
	/evidence=experimental								
	/product="O-succinylbenzoic acid (OSB) CoA ligase"								
	/protein_id="AAAS0402.1"								
	/db_xref="GI:557493"								
	/translation="MORAQLPPIALIEDQVTFAPELFAASKRMAEOLAAHSVRKRG								
	DTAAILLONRAEMVYVAHACPLGVKAVLLTKSTHERLFLQLEDSCSGFLTDSPEE								
	KKEEHLVOTIDVDELMKEAEFEIEIEYQMOMDATLMTLSGTTGPKGVQOQFNGHNN								
	YFSAECRPLRGITREQDAGLCTAALHISGLSLFESVYIKMTIVLHQFSTYECALFYQ								
	HS"								

Query Match	9.1%	Score 108.6;	DB 2;	Length 5932;
Best Local Similarity	49.3%	Pred. NO. 2.1e-22;		
Matches 316;	Conservative 0;	Mismatches 319;	Indels 6;	Gaps 1;

QY	521	TAGCTGACCAGCGGCATACGTATCATGTGTGTTTAAACGCACATGATCATTTGTTATT	580
Db	883	TGGACGATCCACAGCAACGATGTGTTTTTGGCAATTGGAACAAGAGCAAAACCTTTGTCG	942
QY	581	CAGCAACACCGAAGCGGTAGTGGCTATGTCAAGTGGTCAATGCGTACGGCGCGGTCG	640
Db	943	GCGCGTCTCCGGAAGAACTGATCAAAAGAGACGGCGGACACTGTCTATGCTCTCTGTCTGG	1002
QY	641	CTGGGACAAGCCGGCGCGGGAGCGATGGCGCTGACGATATCCGCTTACGGCAAGCGTTGT	700
Db	1003	CAGCGTCCATTAAACGAGGCGGTGAATGGAAGAAGACGACCGCGGATGAGCTTCAATATAT	1062
QY	701	TAGCCAGTCAGAAAAACCGCAFTTGAACATCAATATGTGTGGCAAGTATCAACGACCGCT	760
Db	1063	TAAACGATGAGAAGAAACCTCGCTTGACATGTGATTTGGTAGTATTCATATATGCTT	1122
QY	761	TGCAAGACGTGACGACGTCGCTTAAAGTGGCGCGGCATGCCAAGTTTACTCAAAAATAAGC	820
Db	1123	TTGTATCAACGCTGTTCAAGGTTGAAAAGCCTTGATGGACCTGTTTGTACAAACAMAAA	1182
QY	821	AAGTTACACATTTGTATACACACCAATTACAGGGGAGCATTTGCGGACATTTAAGTGTGACG	880
Db	1183	GCGTTCACACACCTGTTTACGCGGATGTGCGAGACGCTCGCGAGTCTTCGATTTTTCG	1242
QY	881	CGATTGTGGACCGCTGTGCATCCACACGACGACACTGTGGCTGCCACGTAAGCGGCC	940
Db	1243	ACTTAAATTGAAGAGTGTCACCCCACTCCCGCGCTTGGGGGATCACTCAGAAAAAGTG	1302
QY	941	TGTATTACATTTGCGACCCCATGAGAAGACACCTCGTGGCGCTTTTTCACAGTCTTATGGCT	1000
Db	1303	TTGACGTATCAGGAGAGATTGAACCGATGTCCCGCGCTGGTATAGCGGCCCTATCGGCT	1362
QY	1001	ATTTTACCGCAGATPATAGTGGGGAATTTGTGTGGCATCCGTTCCATGTATGTGAATC	1060
Db	1363	GGATTGACAGCCAGAGATPAACGGGGAAATTCGGGTCGCAATTCGTTACAGGCGCTCATCGAG	1422
QY	1061	AAACCCACGACGAGCAACTTATTTGCTGTGGCGGGATGTGGCTGACATCCATCGCATCGCC	1120
Db	1423	GAA-----GCACAGCAAGGCTGTTTGGCGGGGTGGGATTTGTGAGAACACTGCGAACC	1476
QY	1121	AACAGATATGATGAGAACTGGGTTGAATTTGAACCCATG	1161
Db	1477	TCTCGATATGATGAGAAACATGATTAATTTGAAGCCGAT	1517
RESULT	3		
BACMENAOOP			
LOCUS	BACMENAOOP	5946 bp	DNA
DEFINITION	Bacillus subtilis menaquinone operon: menF, menD, menB and menE genes, complete cds.		
ACCESSION	M74538 M74182 M74183		
VERSION	M74538.1 GI:1185287		
KEYWORDS	menaquinone.		
SOURCE	Bacillus subtilis		
ORGANISM	Bacillus subtilis (strain RB1) DNA.		
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group;		
AUTHORS	Bacillus/Staphylococcus group; Bacillus.		
TITLE	1 (bases 1 to 5946)		
	Miller, P., Mueller, J., Hill, K. and Taber, H.		
	Transcriptional regulation of a promoter in the men gene cluster of		
	Bacillus subtilis		
JOURNAL	J. Bacteriol. 170 (6), 2742-2748 (1988)		
MEDLINE	86227858		
REFERENCE	2 (bases 1 to 5946)		
AUTHORS	Driscoll, J.R. and Taber, H.W.		
TITLE	Sequence organization and regulation of the Bacillus subtilis menB operon		
JOURNAL	J. Bacteriol. 174 (15), 5063-5071 (1992)		

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS

J. Bacteriol. 174 (15), 5063-5071 (1992)  
923332443  
3 (sites)  
Thierk, C., Gauss, P., Thernes, C., Groebe, D. R., Gayle, M., Guild, N.,  
Storno, G., d'Abendon-Carafa, Y., Uhlenbeck, O. C., Tinoco, I. Jr.,  
Brody, E. N. and Gold, L.

TITLE CUUCCG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 85 (5), 1364-1368 (1988)  
MEDLINE 88144437  
REFERENCE 4 (bases 1 to 3913)  
AUTHORS Rowland, B., Hill, K., Miller, P., Driscoll, J. and Taber, H.  
TITLE Structural organization of a Bacillus subtilis operon encoding menaquinone biosynthetic enzymes  
JOURNAL Gene 167 (1-2), 105-109 (1995)  
MEDLINE 96144257  
REFERENCE 5 (bases 1 to 5946)  
AUTHORS Rowland, B.  
TITLE Direct Submission  
JOURNAL Submitted (21-OCT-1991) Belinda M. Rowland, Division of Infectious Disease, Wadsworth Center for Laboratories and Research, New York State Department of Health, Albany, NY 12201-2002, USA  
COMMENT On Feb 15, 1996 this sequence version replaced gi:143180.  
FEATURES  
source  
1. .5946  
/organism="Bacillus subtilis"  
/strain="RB1"  
/db\_xref="taxon:1423"  
14. .28  
/note="putative"  
/citation=[1]  
/function="regulatory site for pH regulation of promoter"  
14. .19  
36. .41  
48. .>590  
/note="5' RNA terminus mapped by high-resolution S1 nuclease protection assays (J. Bacteriol. 170, 2742-2748, 1988); exact 3' terminus unclear, but transcripts into menB are undetectable; contains ORF1 to possible ORF5"  
/evidence="experimental"  
130. .134  
143. .1558  
/gene="menB"  
143. .1558  
/gene="menB"  
/note="based on similarity to E. coli EntC, A. hydrophila AmoA, B. subtilis Dhbc"  
/codon\_start=1  
/transl\_table=11  
/product="isochorismate synthase"  
/protein\_id="AAC37013.1"  
/db\_xref="GI:1185288"  
/translation="MTVTQRTFRKEVLAHLKAKENHVAIVSYSGJESIDPLSFFNYGAKKRYTGNRFPMSPDESLITVIGIKFAVOTNKNSERYREVEQMERKKTAFFIYEERLQHSNAGPVILFGSFSFDPCERSGSDHRSQDPFPALMLTMTAEGPILTYNRMVSGGDEAVLEBGLKFAAEFVVPDKQEDQVIAAEIDKDDMLKALETATSQIREKQDKVLAARELLTDPDGIQIEPVAKTLDQOTSYVAIDEGTFVGASPERLIKRDGTVMSCLAGSIRKGVNEEDDRIGLELLEKRLNLEHDIYVGMINAEFSSCEVEKPDGPVLVYKTSVOHLEFTPIVGOLRESASIFDLLEKLPALPGSPOEKVDVARETEPMRGRVYAAPIGWIDODNGEFVAVALRSGLIEGSTARLFGCGCIVDESEIIS EYEETQIKRKPMISALGERR"  
1541. .3201  
/gene="menD"  
1541. .1545  
/gene="menD"  
1555. .3201  
/gene="menD"  
/note="based on similarity to E. coli MenD and enzymatic analysis of mutants (J. Bacteriol. 176, 2648-2653, 1994)"  
/codon\_start=1  
/transl\_table=11  
/product="2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase"  
/protein\_id="AAC37014.1"  
/db\_xref="GI:1185289"  
/translation="MTVNPITHTYIGSFIDEFALSGITDAVCGSRSTPLAVLCANHPDISVHOIDERSAGFALGAKAKORPVLLICTSGTAAANFYPAVVEAHYSRPIIVLTADRPHELARVAGAPQAINQHFLFGNFVKFTFSDALPEESPQMLRTYIRTLPSRAAGAAQ

KRMGCVHVNPLREPLMDPLDSEPPRMRRTGHRVSKTGTOGVDSRESLSDYAEMLAEARKGVTCGEISDADKENITLAKLQYPIIADPISNIRNCYCHKSTYIDAVIDFGLDDELKRLKRPDIYTRGPMRVPKPVFLMLKDDPTIQCIYIDECCGRDPTQASAMHICNASTVAEIEIMAGLITATRRSSSEWLEKMFVNGFHEHLOTTISSEDSFEGNLYRILOH LVPENSLFVGNSMPTIRVDYTFEKKODRPFRIYSNKGANGIDGVSSAGVCEGTAPVTLVIDLSFEYHINGLINLAARKIGIPLYIVLVNDCGJFESFPQASEFTHEDELTGTPGLDFKHAALAYGTYISCPASWDEKFTAVARQTSPOSI"  
3269. .3273  
3284. .3871  
/gene="ORF3"  
3284. .3871  
/gene="ORF3"  
/note="similar to triacylglycerol lipase of P. immobilis"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAC37015.1"  
/db\_xref="GI:143185"  
/translation="MGTVNITVSDGYRAVADGPNASRAVVCILGFTGSKQSWPFLD EMLPDSRLKIDICLGERTDAPLNGKRYSTTRQVSDPLAEIPDLKHLKRYLSMKG RLAYSFAMTYPSGVRHLCLKARRRSGKHLGNGNCGSGTSLILFEYTGSRLLKRG RLSLCPHLRSGWKIFGTGYDRAVCGTIRGLRTV"  
4103. .4117  
/note="sequence motif seen in promoter regions for several B. subtilis genes involved in TCA cycle and electron transport: putative"  
/citation=[2]  
/function="involved in pH regulation of promoter"  
/phenotype="unknown"  
4116. .4121  
/gene="menB"  
/note="menB"  
/note="identified by appropriate spacing from mapped 5' menB RNA and homology to B. subtilis vegetative type promoters"  
/citation=[2]  
4116. .5736  
/gene="menB"  
4139. .4144  
/gene="menB"  
/note="menB"  
/note="identified by correct spacing from a mapped 5' end of menB transcripts"  
/citation=[2]  
4152. .>5736  
/gene="menB"  
/note="menB"  
/note="5' RNA terminus mapped by high resolution primer extension assays.; Exact 3' terminus is unclear but extends at least through menE."  
/citation=[2]  
/evidence="experimental"  
4162. .4166  
/gene="menB"  
/note="menB"  
/note="putative"  
/citation=[2]  
4171. .4956  
/gene="menB"  
/note="menB"  
/note="predicted MW of 28584.4 Daltons, 261 a.a. residues, estimated pI=5.99"  
/citation=[2]  
/codon\_start=1  
/transl\_table=11  
/function="convert OSB-CoA to DHNA in menaquinone biosynthesis"  
/evidence="experimental"  
/product="dihydroxyphenolic acid (DHNA) synthetase"  
/protein\_id="AAC37016.1"  
/db\_xref="GI:143186"  
/translation="MKIPDETNGICIKITINRPEVNAATPPTVAMHINAFADARDQ NVGVIYLAGAGDAFCSDQKVRGKGIVGDDQIPRLNVLDLQRLIRITPRVVAAM SGVAIGGCVHLHIVCDUTIAADNAIFGQVPRKGSFDAGYGGYLARIYGHKRAEIV YLCROYNAOEALDMGLVNTVPLEOLEEETIKWCEMLERKSTALRFLFAANADTDG LAGIOOFAGDAILTLVYTDEAKRGDSFKEKKRPDRCGPRPP"  
4977. .5007  
/gene="menB"

terminator

		/note="contains a loop sequence shown to provide unusual stability [see misc. feature 4975-4982]"
		/citation=[2]
		/evidence=experimental
		4989..4996
	misc_feature	/gene="memB"
		/standard_name='Golden Loop' sequence"
		/note="sequence is very similar to that shown by Therk et al [Proc. Natl. Acad. Sci. U.S.A. 85, 1364-1368 (1988)] to confer stability to RNA stem-loop structures."
		/citation=[3]
		/citation=[2]
		/function="possible stabilization of RNA secondary structure"
		/evidence=experimental
	gene	5000..5946
		/gene="mene"
		5000..5027
	misc_signal	/gene="menB"
		/note="putative"
		/citation=[2]
		/function="recognition sequence for pH regulation of promoter"
		/phenotype="unknown"
	-35_signal	5016..5021
		/gene="menE"
		/citation=[2]
	-10_signal	5038..5043
	Query Match	9.1%; Score 108.6; DB 2; Length 5946;
	Best Local Similarity	49.3%; Pred. No. 2,le-22;
	Matches 316; Conservative	0; Mismatches 319; Indels 6; Gaps 1;
OY	521 TAGCGAGCAGGCGCAATACGTATCATCTGTGGTTTAAACGCANATGATGATTGTTATT 580	
Db	884 TGGACGATGACGAGCAAGCTATGTTTTTGCAATGAACAAGAAGCAAAACCTTTGTGCG 953	
OY	581 CAGCAACACCGAAGCGGTAGTGGCTATGTCAGGTGTCAGATCCGCTACGGCGGGCTGCG 640	
Db	954 GCGCGCTCTCCGAAAGACTGATCAAAGAAGACGGGGCACCTGTCATGTCTTCTGTCGCG 1013	
OY	641 CTGGACAAGCCGCCGGGAGACGATGGCGCTGACGATATCGCGTTAGGCGAAGCGTTGT 700	
Db	1014 CAGGCTCCATTAAACGAGAGCGGTGAATGAAGAAGACGCCGCGGATAGCGCTTGAATTAT 1073	
OY	701 TAGCCAGTCAGAAAACCGCATTTGAACATTCATATGCTGTGGCAGATTCACGACACGCT 760	
Db	1074 TAAACGATGAGAAAAACCTCGCTTGAGCATGATATTGTGTAGTAATTCATAATGCTT 1133	
OY	761 TGCAAGACGTGACGACGTGCTAAAGCTGCGCGGCATCCAAAGTTTACTCAAAAATAAGC 820	
Db	1134 TTGTATCAAGCGCTTCAAGAGTTGAAAAGCCGTGATGAGACCTGTTTGTACAAAACAAAA 1193	
OY	821 AAGTTCAGCACTTTGTACACACCATTAAAGGGGACATTGCGGCACATTTAACTGTGACCG 880	
Db	1194 GCGTTCACACCTCTGTTTAGCGGATTTGTTCGACACACTGCGGAGCTGCTTCGATTTTGG 1253	
OY	881 CGATGTTGAGCGGTTGGATCCAACAGCAGCACTGGGTGGCGTCCACAGCTGAAGCGGCC 940	
Db	1254 ACTTATATTGAGAGCTGCAACCCCCTCCCGGCTTGGGGGATCACCTCGAAAAAGCTG 1313	
OY	941 TGTATTACATTGCGACCCCATGAGAAAGACACCTCTGCGCTGTTGTTTGCAGATCCATTGGCT 1000	
Db	1314 TTGACGTGATCAGGAGATGTGAACCGATGTCCCGGGGTGGATGAGCCCTATCGGCT 1373	
OY	1001 ATTTTACCGCAAGATTAATGCTGGGAATTTGTGTGGCATCCGTTCCATGTAATGGAATC 1060	
Db	1374 GGATTGACAGCCAGGATTAACGGGGAATTCGGGTCGCCATTGCTTCAGGGGCTCATCGAAG 1433	
OY	1061 AAACGCAGCAGCAGACACTTATTTGGTGGGCGCCGGAATGTGGCTACCTCCGATGGGC 1120	
Db	1434 GAA-----GCACACCAAGCGCTGTTTGGCGGGGTGCGATTTGTGGAAGACTCGAACCA 1487	

```

Oy      1121 AACGAGATATGAGAAGACTGGCTGGAATTGCAACCATG 1161
Db      1488 TCTCGGATATGAGAACAACAGATTAAATTGAACCCGATG 1528

RESULT  4
SUB0016/c

LOCUS   BSB0016 215640 bp DNA BCT 26-NOV-1997
DEFINITION Bacillus subtilis complete genome (section 16 of 21): from 2997771 to 3213410.
ACCESSION 299119 AL009126
VERSION 299119.1 GI:2635411
KEYWORDS

SOURCE   Bacillus subtilis.
          Bacillus subtilis
          Bacteria; Firmicutes; Bacillus/Clostridium group;
          Bacillus/Staphylococcus group; Bacillus.

REFERENCE
AUTHORS  Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
          Azavedo,V., Bertero,M.G., Bessieres,P., Boloitin,A., Borchert,S.,
          Boriss,R., Boursier,L., Brans,A., Braun,M., Brignelli,S.C.,
          Biron,S., Brouillette,S., Brusch,C.V., Caldwell,B., Caprano,V.,
          Carter,N.M., Choi,S.K., Codani,J.J., Conerton,I.F., Cummings,N.J.,
          Daniel,R.A., Denizot,F., Devigne,K.M., Dusterhoft,A., Ehrlich,S.D.,
          Emerson,P.T., Enllan,K.D., Errington,J., Fabret,C., Ferrari,E.,
          Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
          Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Goldlighty,E.J.,
          Grandi,G., Guiseppli,G., Guy,B.J., Hega,K., Halech,J., Harwood,C.R.,
          Henaud,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
          Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
          Klier-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
          Konigstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
          Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
          Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
          Mizuno,M., Mostel,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
          Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
          Portetelle,D., Porvollik,S., Prescott,A.M., Prescan,E., Pujic,P.,
          Punelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
          Rivolta,C., Roche,E., Roche,B., Rose,M., Sadate,Y., Sato,T.,
          Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
          Sekowska,A., Seror,S.J., Serro,P., Shin,B.S., Soido,B.,
          Socokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
          Takeuchi,M., Takakoshi,A., Tanaka,T., Terpiltra,P., Tognoni,A.,
          Tosaot,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
          Viari,A., Wanduit,R., Wedler,E., Wedler,H., Weltzenegger,T.,
          Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
          Yoshida,K., Yoshikawa,H.F., Zumbstein,E., Yoshikawa,H. and
          Danchin,A.

TITLE    The complete genome sequence of the gram-positive bacterium,
          Bacillus subtilis

JOURNAL  Nature 390 (6657), 249-256 (1997)
MEDLINE  98044033
REFERENCE
AUTHORS  Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
TITLE    Direct Submission
SUBMITTED (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES
Source   Location/Qualifiers
         1..215640
            /organism="Bacillus subtilis"
            /strain="168"
            /db_xref="taxon:1423"
            complement(8..1001)
            /gene="ytm"
            complement(8..24)
            /gene="ytm"
            complement(99..1001)
            /gene="ytm"
            /function="unknown"
            /codon_start=1

```



[illegible][illegible]

QY	641	CTGGGACAAAGCCGCGGAGCGATGCGCGTACGATATCCGTTAGCGAAGCCTTGT	700
DB	154156	CAGGCTCCATTAAAGAGGCGGATGAGAGACGACCGCGGATAGGCCCTTGATTTAT	154097
QY	701	TAGCCAGTCAGAAAAACCCGATTGAACATCATATGTGCTGGCAAGTATCACACACGCT	760
DB	154096	TAAACGATGAGAAAAACCTGCTTGAGCATGATATTGTGTAGTATTCATTAATGCTT	154037
QY	761	TCCAGAGCTGACGACGTGCTTAAAGGTGCGGCGCATGCTTACTCAAAAAATTAAC	820
DB	154036	TGTGTTCAAAGCTGTTACAGAGTTGAAAACCCGTGATGACCTGTTTGTACAAAACAAAA	153977
QY	821	AAAGTCAGATTGTTACACACCAATTACAGGGAGACATTGCGGCACATTTAAAGTGCACCG	880
DB	153976	CGGTCAGACACTGTTTACGCCGATTGTGCGACAGCTGGCGCAGTGTGCTTGATTTTGG	153917
QY	881	CGATTGTTGACCGCTTGACATCCACACACGACACTGGGTGGCGTCCACGTGAAGCGGCC	940
DB	153916	ACTTATTTGAGAGCTGACCCCACTCCCGCTTGCGGGGATCACCTCAGAAAAAGCTG	153857
QY	941	TGTATTACTTGGCAACCCATGAGACACCTGCGTGTGTTTGTGAGTCTTATTTGCT	1000
DB	153856	TTGACGCTGATCAGGAGATTGAACCGATGTCCCGCGCTGTATGCGGCCCTATCGCT	153797
QY	1001	ATTTTACGCGACATATAGTGGGAATTTGTGTTGGCATCCGTTCCATGTATGTGAATC	1060
DB	153796	GGATTGACAGCCAGGATTAACGGGAATTCGCGTCCCATTTGTCAGGGCTCATCGAAG	153737
QY	1061	AAACCGACGACGACCAACTTATTTGCTGTGTGCGCGGATTTGTGCTGACCTCCGATCGC	1120
DB	153736	GAA-----GCACACGACGAGCGCTTTGCGCGGTATGTGGAACACTCGAACC	153683
QY	1121	AACAGAAATATGAGAACTGGTTGAAATTTGAACCCATG	1161
DB	153682	TCTCGAATATGAGAAACACAGATTAATTAAGCCGATG	153642
RESULT	5		
LOCUS	AF008220	220060 bp	DNA
DEFINITION	Bacillus subtilis rnb-dnaB genomic region.	BCT	04-FEB-1998
ACCESSION	AF008220		
VERSION	AF008220.1	GI:2293135	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS</			

REFERENCE	17 (bases 196487 to 200620)
AUTHORS	Ji,S., De Jesus-Berrios,M. and Sonenshein,A.L.
TITLE	A <i>Bacillus subtilis</i> malate dehydrogenase gene
JOURNAL	<i>J. Bacteriol.</i> 178 (2), 560-563 (1996)
MEDLINE	96134995
REFERENCE	18 (bases 129888 to 132207)
AUTHORS	Varon,D., Brody,M.S. and Price,C.W.
TITLE	<i>Bacillus subtilis</i> operon under the dual control of the general stress transcription factor sigma B and the sporulation transcription factor sigma H
JOURNAL	<i>Mol. Microbiol.</i> 20 (2), 339-350 (1996)
MEDLINE	96310371
REFERENCE	19 (bases 81540 to 91690)
AUTHORS	Bower,S., Perkins,J.B., Yocum,R.R., Howitt,C.L., Rahaim,P. and Pero,J.
TITLE	Cloning, sequencing, and characterization of the <i>Bacillus subtilis</i> biotin biosynthetic operon
JOURNAL	<i>J. Bacteriol.</i> 178 (14), 4122-4130 (1996)
MEDLINE	96312354
REFERENCE	20 (bases 49093 to 51682)
AUTHORS	Yocum,R.R., Perkins,J.B., Howitt,C.L. and Pero,J.
TITLE	Cloning and characterization of the metE gene encoding S-adenosylmethionine synthetase from <i>Bacillus subtilis</i>
JOURNAL	<i>J. Bacteriol.</i> 178 (15), 4604-4610 (1996)
MEDLINE	96343628
REFERENCE	21 (bases 100760 to 102298)
AUTHORS	Kappes,R.M., Kempf,B. and Bremer,E.
TITLE	Three transport systems for the osmoprotectant glycine betaine operate in <i>Bacillus subtilis</i> : characterization of Opud J. <i>Bacteriol.</i> 178 (17), 5071-5079 (1996)
JOURNAL	
MEDLINE	96359357
REFERENCE	22 (bases 1 to 220060)
AUTHORS	Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,S.D.
TITLE	Sequencing and functional annotation of the <i>Bacillus subtilis</i> genes in the 200 kb rnm-dnaB region
JOURNAL	<i>Microbiology</i> 143 (Pt 11), 3431-3441 (1997)
MEDLINE	98048467
REFERENCE	23 (bases 1 to 220060)
AUTHORS	Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,D.
TITLE	Direct Submission
JOURNAL	Submitted (13-JUN-1997) Laboratoire de Genetique Microbienne, Domaine de Valvert, Jouy-en-Josas cedex 78352, France
FEATURES	
Source	Location/Qualifiers 1..220060 /organism="Bacillus subtilis" /db_xref="taxon:1423" 485..2034 /product="16S ribosomal RNA" 2203..5129 /product="23S ribosomal RNA" 5185..5300 /product="5S ribosomal RNA" 5322..5397 /note="codon recognized: GUA" /product="tRNA-Val" 5430..5504 /note="codon recognized: ACA" /product="tRNA-Thr" 5543..5618 /note="codon recognized: AAA" /product="tRNA-Lys" 5629..5715 /note="codon recognized: CUG" /product="tRNA-Leu" 5721..5795 /note="codon recognized: GGC" /product="tRNA-Gly" 5810..5895 /note="codon recognized: UUA" /product="tRNA-Leu" 5905..5981 /note="codon recognized: CGU" /product="tRNA-Arg"

Query Match	Similarity	Score	DB 1:	Length	220060:
<p>Best Local Similarity 49.3%; Pred. No. 3.3e-22;</p> <p>Matches 316; Conservatively 0; Mismatches 319; Indels 6; Gaps 1</p>					
521	TGCGTACGAGCGCATACGATATCATGTGGTTTAAAGCGCATGATGTAATGTTATTT	580	5997. .6073	/note="codon recognized: CCA"	trNA
26151	TGGACGATCAGCAGACACAGCTTGTGTTTGCATTTGAACAGAGGCAAACTTTGTG	26210	6079. .6151	/product="tRNA-Pro"	trNA
581	CAGCAACACCGGAGCGGTAGTGGCTATGTCAGTGGTCAGATCGCTACGGCGCGGTG	640	6172. .6248	/note="codon recognized: AUG"	trNA
26211	GGCGCTCCGGAAAGACTGATCAAAAGAGAGCGGGGCACTGTCATGCTTCTGTCGTG	26270	6251. .6327	/product="tRNA-Met"	trNA
641	CTGGCAGACAGCCGGCGGGGAGCGATGGCGCTGACGATATCCGCTTAGCGCAAGCTTG	700	6334. .6425	/note="codon recognized: UCA"	trNA
26271	CAGGCTCCATTAAACGAGGCGGTGATGAAGAGACAGCCGCGGATGAGCTTGAATTTAT	26330	6443. .6519	/product="tRNA-Ser"	trNA
701	TAGCCAGTCAAGAAAAACCGATTTGAACATCAATATGTCGTGCAAGTATCAGCAGCGCT	760	6531. .6607	/note="codon recognized: AUG"	trNA
26331	TAAAGATGAGAAAAACCGCTTGTGACATGATATTTGTTAGTATGATTCATTAAGCTT	26390		/product="tRNA-Met"	trNA
761	TGCAAGACGTGACGACGTCGCTAAAGTGCGCGGCATGCCAAGTTTACTCAAAAATTAAC	820		/note="codon recognized: GAC"	trNA
26391	TGTATCAAGCTGTTCAGAGGTTGAAAAACCGTATGACCTGTTTGTACAAAACAAAA	26450		/product="tRNA-Asp"	trNA
821	AACTTCAGATTTGTACACACCAATTTACAGGGACATTCGGCAGCATTTAAGTGTACCG	880			
26451	GGCTTCAGACCTGTTTAAAGCCGATGTGCGACAGCTGGCGCAGTCTGCTCGATTTTG	26510			
881	CGATGTTGACCGCTTGCATTCACACACGACACTGGGTGGCGTCCACCGTGAAGCGGCC	940			
26511	ACTTATTTAGAAAGTGCACCCCACTCCCGCTTGGGGGATCACCTCAGGAAAAAGCTG	26570			
941	TGTATTACATTCGACCCATGAGAAAGACACTGCTGGCTTGTTCAGAGTCCCTATTGGCT	1000			
26571	TTCAGCTGATCAGGAGGATTTGAACGATGTCGCCGGGCTGATGATGGCCCTATGGCT	26630			
1001	ATTTCACGCAATATTAAGTGGGGAATTTGTGTTGGCATTCGCTTCATGATGTGATTC	1060			
26631	GGATTCAGACGCAAGGTTAAAGGGGGAATTCGCGCTCCCTGTTGAGGGCTCATCTGAAG	26690			
1061	AAAGCAGGACGACGAACATTTATTTGCTGGTGGCGGGATTTGGGCTGACTCCGATGGCC	1120			
26691	GAA-----GCACAGGAAGGCTGTTGGCCGGGTGGGATTTGTTGGAGACTCGGAACCA	26744			
1121	AACAGAAATATGAGAAACTGGTTGAATTTGAACCCCATG	1161			
26745	TCTCGAATATGAGAAACAGATTTAATTTGAAGCCGATG	26785			

LOCUS	PAX82644	4662 bp	DNA	BCI	08-JAN-1997
DEFINITION	P.aeruginosa pchd, pchc, pchb and pcha genes.				
ACCESSION	X82644				
VERSION	X82644.1 GI:1628425				
KEYWORDS	pchd gene; pchb gene; pchc gene; pchd gene.				
SOURCE	Pseudomonas aeruginosa.				
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group; Pseudomonas.				
REFERENCE	1 (bases 1 to 4662)				
AUTHORS	Serino, L., Reilmann, C., Baur, H., Beyeler, M., Visca, P. and Haas, D.				
TITLE	Structural genes for salicylate biosynthesis from chorismate in Pseudomonas aeruginosa				
JOURNAL	Mol. Gen. Genet. 249 (2), 217-228 (1995)				
MEDLINE	96086939				
REFERENCE	2 (bases 1 to 4662)				
AUTHORS	Serino, L., Reilmann, C., Visca, P., Beyeler, M., Chiesa, V.D. and Haas, D.				
TITLE	Biosynthesis of pycocellin and dihydroaeruginolic acid requires the iron-regulated pchCBA operon in Pseudomonas aeruginosa				
JOURNAL	J. Bacteriol. 179 (1), 248-257 (1997)				
MEDLINE	97136628				
REFERENCE	3 (bases 1 to 4662)				
AUTHORS	Serino, L.				
TITLE	Direct Submission				
JOURNAL	Submitted (14-NOV-1994) L. Serino, University of Lausanne, Lab de Biologie Microbienne, Batiment de Biologie, 1015 Lausanne-Dorigny, SWITZERLAND				
REMARK	Revised by (4)				
REFERENCE	4 (bases 1 to 4662)				
AUTHORS	Serino, L.				
TITLE	Direct Submission				
JOURNAL	Submitted (27-SEP-1996) L. Serino, University of Lausanne, Lab de Biologie Microbienne, Batiment de Biologie, 1015 Lausanne-Dorigny, SWITZERLAND				
COMMENT	On Oct 23, 1996 this sequence version replaced gi:929778.				
FEATURES	Location/Qualifiers				
source	1. 4662				
	/organism="Pseudomonas aeruginosa"				
	/strain="PA01"				
	/db_xref="taxon:287"				
	211. 229				
misc_feature	/note="Fur binding site ('Fur box')"				
gene	join(256..260,265..1905)				
	/gene="pchd"				
RBS	256..260				
	/gene="pchd"				
	265..1905				
	/gene="pchd"				
CDS	/note="similarity to adenylate-forming, activating enzymes"				
	/codon_start=1				
	/transl_table=1				
	/protein_id="CA57966.1"				
	/db_xref="GI:1628426"				
	/db_xref="SPRMBL:P72175"				
	/translation="MTSSPYPSAVDDADPMPAFVRYRLDAGHMODSPFAELATSA AHPRRLACDDDRSLSTADLQRCRRRLAAGLRQGLAHGDTTVVHLPLNGIAFVTCRA LFQGVRFVIALPFRHROHETSGCFRFAEAKAYICAEIDGDEPMPRLRELASACRRA ALIHGEAEVPLALAPLYOALDEDCARAEADIAFCPLSGTTPPLIPRHHRELY NVRASAEVCGFDEHTVYLTGIPMAHNFLLCCPGVIGTLASGAVVQADPEHCFL IAREVHTHTALVPLAMLMADQESRADLSLLQVGSRLGSSAORVEPYLGGC LOOVLGMAEGLICTRLDDPRLHTOGRLSPDDRVYVDAGSRVGGGVEGELTV RGPYTRIGYRLPREHNAKASADGFYTGDSVSDKRGYLYVEBRDDQINRGSEKTA AEEVENLLIAPHOVHDAIVVAMPDSLGEETCAVITRPPASALKLKYLIHACGLAA FKVPDRIELVAPFQGTIGIKISKRRLERLRLELARA"				
gene	1890..1894				
	/gene="pchc"				
RBS	1890..1894				
	/gene="pchc"				
gene	1902..2654				
	/gene="pchc"				
CDS	1902..2654				

gene	/gene="pchc"		/note="similarity to thioesterases"		
RBS	/codon_start=1				
gene	/transl_table=1				
CDS	/protein_id="CA57967.1"				
	/db_xref="GI:1628427"				
	/db_xref="SPRMBL:P72176"				
	/translation="MSAAWVRPRLTTPMRLRLACFPHAGSASFPSRSPRLPPDID LLAQYPEREDRENEAPTRLEDDALALRFPADAPLALFHSIGALATYALR LETPAURLHFLVSAHPAPHRQSGALRHGDEALLEDVRRQGSLELDEADLRALF PILRADYQALETTYRAQPIALACALDVLGHEDEVSAAEQASDSRTPARLRPP GGHFYLSGGRADVTEHLRLRLHPDALSREVA"				
	2641..2646				
	/gene="pchb"				
	2641..2646				
	/gene="pchb"				
	2651..2959				
	/gene="pchb"				
	2651..2959				
	/gene="pchb"				
	/function="conversion of chorismate to salicylate (together with pchA)"				
	/codon_start=1				
	/transl_table=1				
	/protein_id="CA57968.1"				
	/db_xref="GI:929779"				
	/db_xref="SWISS-PROT:Q51507"				
	/translation="MKTPEDCTGLADIREAIDRIDDIYQALGRMDYKKAASRFA SEAAIPAPERVAAMLPERRARAEENGDAPEVEGLAQIILHWYIAEQIKWROTRGAA"				
gene	2956..4386				
CDS	/gene="pchA"				
	2956..4386				
	/gene="pchA"				
	/function="conversion of chorismate to salicylate (together with pchB)"				
	/codon_start=1				
	/transl_table=1				
	/protein_id="CA57969.1"				
	/db_xref="GI:929780"				
	/db_xref="SWISS-PROT:Q51508"				
	/translation="MSRLAPLSOCILALRGTFERRAIGQAOALDRPVLVAASFETPID PLQVFGAMDROTPCLYOEPLAFPMGCALELOGHGEORFARI EEWOLLCADAVY EGPLAPRLGGFRPDRPGPREHMOARADASIMAGITVIRREGRRVVLCOHLAKPG DALALAAHGSALILRLRQPARRRSPGTPAGACQASOERRQMAKYSDAYSVRQGR FGKVVLAHQAPRLGDIEMOVNIEHLRLQHADADLFACRKNACFLTASPRVLRISA GEALTHALAGTIARGDAQEDBARLOLDSANKRHHQIUVENIRTLALPEFSFVLEI PDAPGLRLARVOHLNTPIRARLADAGGILRLLOALHPTPAVGCVPRSAADLYTROHE GMDRGWYAAPLPGMDGEGNGDFIVALSALLTPRGYLVFAGCGLVGDSBPRAHEVRETC LKTSAMREALSITGIDPEVLQRGVA"				
BASE COUNT	714 a	1719 c	1560 g	669 t	
ORIGIN					
Query Match	8.6%	Score 103.4;	DB 2;	Length 4662;	
Best Local Similarity	47.5%;	Pred. No. 8.7e-21;			
Matches 345;	Conservative 0;	Mismatches 376;	Indels 6;	Gaps 1;	
QY	441	CTTAGCCAAATCGCTTTTGGTGGCAGACACACCTCGACGTTATCCGACACGTTACGACT	500		
DB	3609	CTTCGGCGAAGTGTGCTGGCCCGACCCCGCGGCTCTCGGCGACATGCAACCCGCG	3668		
QY	501	GGCACAAATTATTCGTGTGCTTACGTGACGAGCGAATCGATCATGTGGTTTAAAGCG	560		
DB	3669	GCAGTCATTCGAACACCTGCGCTTCGACACATGCGACGCCCGACGCTGTTCGCTTCGCGC	3728		
QY	561	ACATGATCAATTTGTTATTTATTCAGCAACACCGGAAGGTTAGTGCTATGTCAAGTGCTCA	620		
DB	3729	CGGGAAGCCCTGCTTCCTCGGCGCGCTCCCGGAAGCGCTGTGTCGCGCATTCGCGCGGGGA	3788		
QY	621	GATGCGTACGGCGGCGGTGCTGGGACAAAGCGGCGCGGAGCATGGCGCTGACGATAT	680		
DB	3789	GGCACCTACCCATCCCTGCGCGGGGACCATGCGCCCGCGCGCATGGCCAGGAAGATGC	3848		

QY	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS
QY 681	CGCGTTAGCGCAAGCGTTGTTATGACCGAGTACAGAAAACCGCATTTGAACATCATATGTCG				
Db 3849	CGCGCTCGGACAAAGCCCTCTGCTGGACAGGCCCAAGGACAGCAGCAACCAATTGTTGGT				
QY 741	GCGAAGTATCAAGACACGCTTGGCAAGACGTGACGACGTGCTAAAGGTGCCGCGATCGC				
Db 3909	GGAGGCGATCGGTACGGCCCTGGAAACCTTACGAGGAGTGGTGGAAATTCGCCGATGGCGC				
QY 801	AAGTTTACTCAAAAATTAAGCAAGTTCACGATTTTGTACACACCAATTTACAGGGGACATTTC				
Db 3969	CGGCGCTGGAACGACATGGGGCGGAGTCCACACCTGTAACACGCGAATCCGCGCGCTCGC				
QY 861	GGCACATTTAAGTGTGACCGGCGATGTTGACCGCTGGATCCCAACACAGCAGCTGGGTGG				
Db 4029	TGACGCGAGCGCGCATCTCGCGGCTCTCTCAAAAGCTGTGATCCGACCCCGCGTGGCGG				
QY 921	CGTCCACAGTGAAGCGCGCTGTATTTACATTTGCGACCCATGGAAGACACCTCTGCGCTT				
Db 4089	CTACCCACGACGAGCGCGGCTGGACTACATCCGACGACGAAAGGATGAGACCGCGGCGT				
QY 981	GTTTGCAGGTCTCTATTTGGCTATTTTACCGCAGATTAATGTGGGGAATTTGTGTTGCGAT				
Db 4149	GTACGCGCGCGCGCTGCGCTGCGACGGGCAAGGCAACGCGCATTTCTGTGTGCGCT				
QY 1041	CCGTTCCATGTATGTGATGATCAAAAGCAGCAGACGACACATTTAATTTGGTGGGCGCGGAT				
Db 4209	GCGCTCG-----GCCCTGCTCACGCCCGCGCGGCGGCTCACTGTTCCGCGCTGCGGTCT				
QY 1101	TGTGGCTACTCTCCATGCGCAACAAGAAATATGTAAGAACTGGGTTGAATTTGAACCCAT				
Db 4263	GGTAGCGCATTTTCGAAACCGCGCCACAGAGTATTCGCGAACTCTTAAGCTCAGTGCAT				
QY 1161	GCGCGAA 1167				
Db 4323	GCGGGA 4329				
RESULT 7					
LOCUS	D90911	143051 bp	DNA	BCF	07-FEB-1999
DEFINITION	Synechocystis sp. PCC6803 complete genome, 13/27, 1576593-1719643.				
ACCESSION	D90911	AB001338			
VERSION	D90911.1	GI:1653083			
KEYWORDS	(p)ppcp-3-pyrophosphohydrolase; 3-dehydroquininate synthase; 3-isopropylmalate dehydrogenase; ATP synthase b subunit; ATP synthase c subunit; D-alanyl-D-alanine carboxypeptidase; D-isomer specific 2-hydroxyacid dehydrogenase family; GTP-binding protein HflX; HgkK; Mannosyltransferase B; Na <sup>+</sup> -ATPase subunit J; P3 protein; Plax protein; S-adenosylhomocysteine hydrolase; acriflavin resistance protein; adenine phosphoribosyltransferase; apolipoprotein N-acyltransferase; beta ketacyl-acyl carrier protein synthase; beta-ketoacyl-acyl carrier protein synthase III; cation-transporting ATPase PacL; ceramide glucosyltransferase; cytochrome c oxidase subunit II; drdp-glucose 4-6-dehydratase; global nitrogen regulator; glucosyltransferase; hydrogenase large subunit; hydrogenase small subunit; hydrogenase subunit; inorganic carbon transport protein; isochromate synthase; leucine aminopeptidase; menaquinone biosynthesis protein; pentose-5-phosphate-3-epimerase; photosystem I subunit III; photosystem I subunit IX; polysialic acid transport ATP-binding protein KpsrI; polysialic acid transport protein KpsM; porphobilinogen synthase; potential NAD-reducing hydrogenase subunit; processing protease; proline-tRNA ligase; protease HnbB; protoporphyrinogen oxidase; regulatory components of sensory transduction system; resolvase; sensory transduction histidine kinase; short-chain alcohol dehydrogenase family; sialoglycoprotease; spore coat polysaccharide biosynthesis protein SpA; spore coat polysaccharide biosynthesis protein SpC; superoxide dismutase; tRNA delta-2-isopentenylpyrophosphate (IPP) transferase; tRNA-gly; tRNA-phe; transposase.				
SOURCE	Synechocystis sp. (strain:PCC6803) DNA.				
ORGANISM	Synechocystis sp.				

REFERENCE	AUTHORS	TITLE
JOURNAL MEDLINE REFERENCE AUTHORS JOURNAL TITLE	Bacteria; Cyanobacteria; Chroococcales; Synechocystis. 1 (sites)	Kaneho,T., Sato,S., Kotani,H., Tanaka,A., Asanizu,E., Nakamura,Y., Miyajima,N., Hirosewa,M., Sugiyura,M., Sasamoto,S., Kimura,T., Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Nartuo,K., Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A., Yamada,M., Yasuda,M. and Tabata,S.
COMMENT FEATURES SOURCE	Synechocystis sp. strain PCC6803. IT. Sequence determination of the entire genome and assignment of potential protein-coding regions DNA Res. 3 (3), 109-136 (1996) 2 (bases 1 to 143051) Tabata,S. Direct Submission Submitted (28-JUN-1996) to the DDBJ/EMBL/Genbank databases. Satoshi Tabata, Kazusa DNA Research Institute, Laboratory of Gene Structure 2; 1523-3, Yanauchino, Kisarazu, Chiba 292, Japan (E-mail:tabata@kazusa.or.jp, Tel:+81-438-52-3933, Fax:+81-438-52-3934) Similarity search of the ORFs and Genemark analysis. Location/Qualifiers 1. 143051 /organism="Synechocystis sp." /strain="PCC6803" /db_xref="taxon:1143" 173. 1960 /note="ORF_ID:sir2105" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAA18000.1" /db_xref="GI:1653084" /translation="MARLTALFRYGYAFILPAIAFLAGIAGLVGVNMSKYLIGLM AIVAPFIWMWMLMORSOGFLGRSSOASANYASNLVALITLTIINAIYAPVAM DIENQNFRTIAROSQSIVAELEOPAKWIFIDDPASEKDRLLNTORYFSQSFEPH IPDYVLQVGHELPLANNEPRGISTAVNALNGVTQPLNVVIDGGIPEDADAVYING PTKRFLAPEATIAKQYDQDGGLNLYMSPADGLBDLEBPQVSFPPRLVIDLSGAG GVGLDPTVPILFNRYGNHPITOKLOGAIALTEPLVPATKEPHIOATTLVEASDLMY ATOLISFDLPNPETDLGPDDVAVALTRGKESVSKEMETGNMPREINPSPP VSRSPSTGCTGTNPPEAMPENDSNQAQADRESPEAGDEDSNDQAQTQQNGDELP AKMVVGSAAFATWAMEVNAENVGDIPLNSVQMAENVDPPLANRAKDPTDRINLGVR RAYLIGWLTVIVPLIGLIGLAFTWNRQR"	
CDS	gene	/gene="apt" complement(2133..2651) /gene="apt" /note="ORF_ID:sll1430" /codon_start=1 /transl_table=11 /product="adenine phosphoribosyltransferase" /protein_id="BAA18001.1" /db_xref="GI:1653085" /translation="MDLKALIRDIPDFPKPGINERDITTLINSBEGLRYTIDSIVEOC ESGLVDHYVMGESRGFLFGMPLAYOMNAGFTIPVRPGKLPAVHVHEVDLEGKDS LEIHODVAHHRRHLIVDDLITATGSTAKATAELLTKGCEVLGAFFIIELALNGRC LPDLPIITSLEY"
CDS	gene	complement(3032..3946) /note="ORF_ID:sll1429" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAA18002.1" /db_xref="GI:1653086" /translation="MIIVLPMILGAGATATFWGFKNIKNAMEDEKKKEEKRRKRE ARTNEKAIMDAVERYGEMLVFEKERISOEMNOVYSKILSEDRKNKTINLEDIEHLER YIIFSLTYISMETVEDSOSRLIEDHYDTGICKEKENNEKRNKKMKWESSNYINS HIIFSLTYITLDIKKECKEOKLIESHGSDIKMGCVDEKATILHHTPMGYRSGTLOOT QDGCADIILMKDKRKITYOKRSASVSAIDEILGAKGYTDADIAIVYTNNSYTEN AKELAYNKVELWDRLKILEFTLESRI"
CDS	gene	complement(3944..4209)



LOCUS	ECOEINTC	1659 bp	DNA	BCF	25-JAN-1994
DEFINITION	E.coli isochorismate synthase (entc) gene, complete cds.				
ACCESSION	M24142				
VERSION	M24142.1				
KEYWORDS	GI:341117				
SOURCE	entc gene; isochorismate synthase.				
ORGANISM	Escherichia coli.				
REFERENCE	Escherichia coli				
AUTHORS	Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.				
TITLE	1 (bases 1 to 1659)				
JOURNAL	Ozenberger,B.A., Brickman,T.J. and McIntosh,M.A.				
MEDLINE	Nucleotide sequence of Escherichia coli isochorismate synthase gene entc and evolutionary relationship of isochorismate synthetase and other chorisimate-utilizing enzymes				
FEATURES	J. Bacteriol. 171 (2), 775-783 (1989)				
source	89123153				
gene	Location/Qualifiers				
CDS	1..1659				
	/organism="Escherichia coli"				
	/strain="K12"				
	/db_xref="taxon:562"				
	380..1555				
	/gene="entc"				
	380..1555				
	/gene="entc"				
	/EC_number="5.4.99.6"				
	/codon_start=1				
	/product="isochorismate synthase"				
	/protein_id="AA16100.1"				
	/db_xref="GI:450376"				
	/translation="MDTSLAEVVOQTMTATLAPNRFEFMSPYRSFTTSGCFARDEPAVNGSDPSFQOKLALFADAKQGIKPNVWGAIPDPQPSLYIPESQSFROEKQASRRPTSOGLNLVVERQAIPEQTPFEQWVARAAATATADVKVLSLIDITDPAADISQVLEIRLIQNPVSNFHVLAAGVILGASPELLRKDERFSYPLIGSARAPDQVDLDRFAGNRILASEKRDHEHETVQAMKEVLRSSSELHVPSPDILPTLMHLAPFEKANSQENALILACLIHPTPLSLGFPFHQAQVYALEPFPRLFGIVGNCDSRNGMVVYTIKAKIREKMOVRIFAAGATVPASSPLGEMRETKVLTMLNFGH"				
BASE COUNT	379 a 443 c 454 g 383 t				
ORIGIN					
Query Match	6.3%; Score 76; DB 2; Length 1659;				
Best Local Similarity	46.6%; Pred. No. 2.8e-12;				
Matches 284; Conservative	0; Mismatches 320; Indels 6; Gaps 1;				
563	ATGATGAATTCGTTATTTACAGACACCGGAAACGGTTATGCGTATGTCAGGTGTCGA	622			
Db 939	ATGTGTGGCTCTCTGCTGGGGGCGACCCGGAACTGCTCTACGTAAAGACGGCGCTT	998			
QY 623	TCGCTACGGCGCGGCTCCTGTGGAGCAACCGCGCGGAGCGGATGGCGTACGATATCG	682			
Db 999	TTAGCTCCATTCCTGTTACCGGTTCCGGGCGCTGCACCGCGATGAATGCTCATCGCG	1058			
QY 683	CGTTAGCGGAACGCTTTGACCACTCAGAAAAACCGCATTTGAACATCATATATGCTGG	742			
Db 1059	AAGCAGGTAAATCGTCTGCTGGCGTCAAGAAAAAGATCCGCAATGAACATGAACGTG	1118			
QY 743	CAGATATACGACACACGCTTGCAGAACGAGACAGAGCTGCTTAAAGSTGGCGGCAAT	802			
Db 1119	AGCGCATGAAGAGAGTACTGCGGCAAGCGACAGTGAATTTACACGCTCTTCTTCAC	1178			
QY 803	GTTTACTCAAAAATTAAGCAACTTTCAGCAATTTGTACACACCAATTAACAGGGACAT	862			
Db 1179	AGCTGATATACACGCGGACGCTGTGGCATCTTCGCAACTCCCTTTGAAGTAAAGGAT	1238			
QY 863	CACATTTATAGTGTGACCGCATTTGTTGACCGCTTGATCCAAACACAGCACTGGTGG	922			
Db 1239	CGCAGAAAGACATGTGACTGTGGCTTCTGCTGATCCGACCCCGCGCTGACCGGTT	1298			
QY 923	TCCCAAGGAGGCGCCCTGTATTAATCAATTTGGAGACCCATGAGAAGACACTGTTGT	982			
Db 1299	TCCTCCATTCAGGCGCGACCCAGGTTATTTGGTGTGAATCTGGAACCGTTTCGACG	1358			

OY	963	TTGCAGGTCATTGGCTATTTTACCCGACAGATAAAGTGCGGAATTTGGTGGCATCC	1042
Db	1359	TTGGGGGCAATGTGGGTGTGTGTGACACGAGAAGTAACGCCAAATGGGTGGTACCATCC	1418
OY	1043	GTTCCATGTATGTGAATCAACACGCGACGACGACAATTTATTTCTGTGTGGCGGATGG	1102
Db	1419	GCTGGCGCAAGCTGGCGGAAATTCAGTGGCG-----TCTGTTTCCCGAGCGGGGATTTG	1472
OY	1103	TGGCTACTCCGATCCGACACAGATAATGAAGAACTGGGTGTGAATTTGAACCATGCG	1162
Db	1473	TGCCGCGCTTCGACACCGCTTGGGTGAGTGGCGGAAACAGCGCTCAAACCTTTCTACCATGT	1532
OY	1163	GGCAATTGTT 1172	
Db	1533	TGAACGTTTT 1542	
RESULT	9		
AEO00165			
LOCUS	AE000165	12003 bp DNA BCT	12-NOV-1998
DEFINITION	Escherichia coli K-12 MG1655 section 55 of 400 of the complete genome.		
ACCESSION	AE000165 U00096		
VERSION	AE000165.1 GI:1786808		
KEYWORDS			
SOURCE	Escherichia coli.		
ORGANISM	Escherichia coli Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.		
REFERENCE	1 (bases 1 to 12003) Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mu,B. and Shao,Y. The complete genome sequence of Escherichia coli K-12 Science 277 (5331), 1453-1474 (1997)		
JOURNAL	9/42617		
MEDLINE	2 (bases 1 to 12003) Blattner,F.R.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459		
JOURNAL	3 (bases 1 to 12003) Blattner,F.R.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459		
JOURNAL	4 (bases 1 to 12003) Plunkett,G. III.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA		
JOURNAL	This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K-12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky. Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: mark@amer.gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli		
COMMENT			

Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). \*\*\* The E. coli K-12 sequence and its annotations are periodically updated; this is version M4. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

source  
1. .12003  
/organism="Escherichia coli"  
/strain="K-12"  
/sub\_strain="MG1655"  
/db\_xref="taxon:562"  
complement(<1..22)  
/note="boxc"  
complement(52..74)  
/note="central position to predicted promoter:  
-181.5 genetic evidence for the site"  
/bound\_moiety="Fur predicted site"  
126..154  
/note="factor Sigma70; predicted +1 start at 624051"  
complement(138..150)  
/note="central position to predicted promoter: -272.5"  
/bound\_moiety="Ada predicted site"  
173..201  
/note="factor Sigma70; predicted +1 start at 624098"  
218..1393  
/gene="entC"  
/note="b0593"  
218..1393  
/gene="entC"  
/EC\_number="5.4.99.6"  
/function="enzyme; Biosynthesis of cofactors, carriers:  
Enterochelin"  
/note="o391; 100 pct identical to ENTC\_ECOLI SW: P10377"  
/codon\_start=1  
/transl\_table=1  
/product="isochorismate hydroxymutase 2, enterochelin  
biosynthesis"  
/protein\_id="AAC73694.1"  
/db\_xref="GI:1786809"  
/db\_xref="PID:g1786809"  
/translation="WDTSLABEVQQTAMTAPNRFPMSPRSTTSGCFARDEPAY  
NGSDPSDFQOKLALFADAKAQGIKNPVWGAIPDPQPSLYIPESWQSFROEK  
QASARRTFQOSLNVVEROAIPEQTPEOMVAAALATPQVDKVLISRLDITTA  
AIDSGVLEERLIKONPVSYNFHVPLADGVYLGASPELLLRKDGERSPIPLASARR  
QDPEDLREAGNRLASEKDRHEHELVTOAMKEVLRERSSSLHVPSSPOLITPTLMH  
LATPECKANSQENALTLACLHPRTALSGFRHOAAVOYIAELEPPRELFGIVGNC  
DSGNGEMVYVTRAKLRKNOVRLFAGAGIVPASSPLGEMRETGVKLTMLNVGLH  
1403..3013  
/gene="entE"  
/note="b0594"  
1403..3013  
/gene="entE"  
/EC\_number="6.3.2.-"  
/function="enzyme; Biosynthesis of cofactors, carriers:  
Enterochelin"  
/note="o536; 98 pct identical to ENTE\_ECOLI SW: P10378"  
/codon\_start=1  
/transl\_table=1  
/product="2,3-dihydroxybenzoate-AMP ligase"  
/protein\_id="AAC73695.1"  
/db\_xref="GI:1786810"  
/db\_xref="PID:g1786810"  
/translation="MSIPPTRMPEEFARRYREKGYMODLPITDITLRHAASDAVID  
GEROLSYRELNOADNLACSLRGGIKRGETALYOLQNVAEITYTFPALLGLGAPIL  
ALFSDHSELNATASQTEPALILNDROHALFSGDFNTFTFTHSSIRVOLLDSGE  
HNLQDAINHPAEDFTATPSPADEVAVYFQLSGCTGTGPKLIPRTINDYVYSRRSVEIC  
QFTQOTRYLCAIPAAHNVMSSPGLGVFLAGIVVLAADPSATLCFPLLEKHQVNVY

promoter  
/EC\_number="3.3.2.1"  
/function="enzyme; Biosynthesis of cofactors, carriers:  
Enterochelin"  
/note="o285; 100 pct identical to ENTB\_ECOLI SW: P15048"  
/codon\_start=1  
/transl\_table=1  
/product="2,3-dihydro-2,3-dihydroxybenzoate synthetase,  
isochorismatase"  
/protein\_id="AAC73696.1"  
/db\_xref="GI:1786811"  
/db\_xref="PID:g1786811"  
/translation="MAIPKIQAYALPESHDIPOKRVDAFEPORALLIHDMQFVVS  
FMGNCPMQEVIANIALRDYCKOHNPVYTAOPKQSDERALLNDMGPGILTRS  
PEQCKVYDRLPDADDPYLVKWSAFRSPLEOMLKRSNGNOLITGVVNHICMT  
ATDAFMRIKRFVADALADFSRDEHLSLKYVAGRSRVYNTTELLPAPIPASKAL  
REVILPLEDSDEPDDNLIDYGLDSYRMALAAWRKRVHGDIDFVLANPITDAW  
WKLISREYK"  
3884..4630  
/gene="entA"  
/note="b0596"  
3884..4630  
/gene="entA"  
/EC\_number="1.3.1.28"  
/function="enzyme; Biosynthesis of cofactors, carriers:  
Enterochelin"  
/note="o248; 100 pct identical to ENTA\_ECOLI SW: P15047"  
/codon\_start=1  
/transl\_table=1  
/product="2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase,  
enterochelin biosynthesis"  
/protein\_id="AAC73697.1"  
/db\_xref="GI:1786812"  
/db\_xref="PID:g1786812"  
/translation="WDFSGKRWVWTGAGKIGIVTALAFVAGAKKVTGFDGAFYQEOY  
PFATVMDVADAAVOAOCORLAETTERLDALVNAAGILRAGADOLSKEDMOQTFV  
NVGAFNLFQOTMFORBOGALITVASDAAHNPRIQMSAYGASKAKSLASVGL  
ELASGVACNVSGSTPTDMQRTLWSDDAEQRIRGFGQFKGIPKGIARPOEI  
ANTLIFASDIASHIITLDIYVDSGTIGA"  
4488..4516  
/gene="entA"  
/note="factor Sigma70; predicted +1 start at 628413"  
4633..5046  
/gene="ybdB"  
/note="b0597"  
4633..5046  
/gene="ybdB"  
/function="orf; Unknown"  
/note="o137; 100 pct identical to YBD\_B\_ECOLI SW: P15050"  
/codon\_start=1  
/transl\_table=1  
/product="orf; hypothetical protein"  
/protein\_id="AAC73698.1"  
/db\_xref="GI:1786813"  
/db\_xref="PID:g1786813"  
/translation="MIMKRHLTDELNATSDNTWVAHLGIYVTRIGDVLPAEMPVDT  
PTHOPFGILHGSAALAEPTGSMAGFMPRTGGOCYVGTETLNAHHNRPVSGKXGVC  
OPLHIGNONOSWEIVFEDGGRCTCTGLTAYLG"  
5098..5107  
/note="central position to predicted promoter: -66"  
/bound\_moiety="MetJ predicted site"

FEATURES

protein\_bind  
1. .12003  
/organism="Escherichia coli"  
/strain="K-12"  
/sub\_strain="MG1655"  
/db\_xref="taxon:562"  
complement(<1..22)  
/note="boxc"  
complement(52..74)  
/note="central position to predicted promoter:  
-181.5 genetic evidence for the site"  
/bound\_moiety="Fur predicted site"  
126..154  
/note="factor Sigma70; predicted +1 start at 624051"  
complement(138..150)  
/note="central position to predicted promoter: -272.5"  
/bound\_moiety="Ada predicted site"  
173..201  
/note="factor Sigma70; predicted +1 start at 624098"  
218..1393  
/gene="entC"  
/note="b0593"  
218..1393  
/gene="entC"  
/EC\_number="5.4.99.6"  
/function="enzyme; Biosynthesis of cofactors, carriers:  
Enterochelin"  
/note="o391; 100 pct identical to ENTC\_ECOLI SW: P10377"  
/codon\_start=1  
/transl\_table=1  
/product="isochorismate hydroxymutase 2, enterochelin  
biosynthesis"  
/protein\_id="AAC73694.1"  
/db\_xref="GI:1786809"  
/db\_xref="PID:g1786809"  
/translation="WDTSLABEVQQTAMTAPNRFPMSPRSTTSGCFARDEPAY  
NGSDPSDFQOKLALFADAKAQGIKNPVWGAIPDPQPSLYIPESWQSFROEK  
QASARRTFQOSLNVVEROAIPEQTPEOMVAAALATPQVDKVLISRLDITTA  
AIDSGVLEERLIKONPVSYNFHVPLADGVYLGASPELLLRKDGERSPIPLASARR  
QDPEDLREAGNRLASEKDRHEHELVTOAMKEVLRERSSSLHVPSSPOLITPTLMH  
LATPECKANSQENALTLACLHPRTALSGFRHOAAVOYIAELEPPRELFGIVGNC  
DSGNGEMVYVTRAKLRKNOVRLFAGAGIVPASSPLGEMRETGVKLTMLNVGLH  
1403..3013  
/gene="entE"  
/note="b0594"  
1403..3013  
/gene="entE"  
/EC\_number="6.3.2.-"  
/function="enzyme; Biosynthesis of cofactors, carriers:  
Enterochelin"  
/note="o536; 98 pct identical to ENTE\_ECOLI SW: P10378"  
/codon\_start=1  
/transl\_table=1  
/product="2,3-dihydroxybenzoate-AMP ligase"  
/protein\_id="AAC73695.1"  
/db\_xref="GI:1786810"  
/db\_xref="PID:g1786810"  
/translation="MSIPPTRMPEEFARRYREKGYMODLPITDITLRHAASDAVID  
GEROLSYRELNOADNLACSLRGGIKRGETALYOLQNVAEITYTFPALLGLGAPIL  
ALFSDHSELNATASQTEPALILNDROHALFSGDFNTFTFTHSSIRVOLLDSGE  
HNLQDAINHPAEDFTATPSPADEVAVYFQLSGCTGTGPKLIPRTINDYVYSRRSVEIC  
QFTQOTRYLCAIPAAHNVMSSPGLGVFLAGIVVLAADPSATLCFPLLEKHQVNVY





```
misc_feature
protein_id="BAB40732.1"
/db_xref="GI:1778449"
/translation="WQLQPMSHFAPGLSVLPVKNRPKGLARKNGIYASITINEVRMRRLVFCEFCGLMSSSGMAVDPLGTINIMLHCNVDFSCVTADIDKTVDGRNPTVLNLADGTAIVLFPSLRLEGCPCPGSVAIIFTGPSADTLLMLDDPAAQOVAIILRNSDRSRIALSPCTSEEDVANGNTLVNFANYRALASGVAPAKADAIFEMINYN"4315..6834
/feature="previously sequenced by another group and can be found in Genbank Accession Number X51862"
complement(4479..5174)
/gene="fimZ"
/complement(4479..5174)
/gene="fimZ"
/codon_start=1
/transl_table=11
/product="fimbriae 2 protein"
/protein_id="BAB40733.1"
/db_xref="GI:1778450"
/translation="MYOSTSRPNKLKSAPMRLYTMKPSTIIMDTPIIRMSIEVILLQNKSELQIVLKTDYRITFDYIKTRPVLLTIMDDLPETDSFIYLKRKIQOSVKVULFSLSSECFYAGRAIQAANGVSKCNDQNDIFAYOVAILSGYFFPESELTNYKSNKCSNSTSYAVTSNREVTLIRYLVSLGSLSNKKEIADKILLSNKTYSAHKSNITYKGILASHIIVELIYAKYEI"5282..7550
/misc_feature
note="previously sequenced by another group and can be found in Genbank Accession Numbers M27155 and M31074"
product="crRNA-Arg"
5440..24000
/feature="lambda-like prophage sequences; exact boundaries of prophage unsure: between 5440-5446 and 23700-24312"
complement(5446..6609)
/gene="intd"
/complement(5446..6609)
/gene="intd"
/note="prophage DUp12 integrase"
/codon_start=1
/transl_table=11
/product="DUp12 integrase"
/protein_id="BAB40734.1"
/db_xref="GI:1778451"
/translation="MSLFRRNDEIWAYSIPLGGKRIKESICTKKROAELOHDKRAELMRYEKLDLPVYTEEACLNREKADKSLSDSKSTEFMEHFEGIRIKDOSIAKITSAVSRRNNRKTEIKWKOKVOAATIRKKELPVEPKVSTOTAKHLAMKALLRRAENDMWAKAPYIKIPAVNRKRVWLKEEEKRLIDECPKLSVGFALATGILRCNAIINLMOODIKOARVANAPEBSKSNRAIGVALNDTACKVLRDQIGKHKKHEVHTAAAKRAGGSTPAVRKMRIDSKTSMISACSACRAGICEPRFDLHHTWASMLIOGVSLVLOMGEMGESIEVMERRAHLAPHNLTPEHAKKIDDIFGDVPNMNSHSEIMEDIKKA"6603..7163
CDs
note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB40735.1"
/db_xref="GI:1778452"
/translation="MTSFRLCYAITRAHLDMSORRLKRNAGDLPCCEYPGYREDASSFLPYRKTNESCCLVHFDELRLHLVKTIRHVLFIPLNHDIMQATFLAIRIVGPCVISFSRHAFVLCGHVRESFYCLETTPEEPFEIEHGISTSIAGSVACIMOSLIGNRRIRINGMLTPCTRTGHPRMVLTTRYNGARLPMN"7408..8665
misc_feature
join(6717..7407,8666..9383)
/note="previously sequenced by another group and can be found in Genbank Accession Number D12596"
7106..7318
CDs
note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB40736.1"
/db_xref="GI:1778453"
/translation="WRANALADAE LRKADELTCMTARINRGETIPEYKOALPYVMGGPUNRAOLAKLAETIKAFGLKGSV"7408..8665
source
organism="Escherichia coli"
insertion_seg="IS3"
/db_xref="taxon:562"
```

Query Match	Best Local Similarity	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
563 ATGATGAATTGTTTATTTTCAGCAACACCGGGAAGGTTAGTGGCTATGTCAGGTGGTCAGA	46.68	46.68	284	0	320	6	1
66075 ATGTTGGGGCTCTGCTGGGGGGCCAGCCGGAAGTCTGTAAGTAAAGCGCGAGCGTT	46.68	46.68	284	0	320	6	1
623 TCGCTACGGCGGGCTGCTGGGACAAAGCGCGCGGAGCGAGTGGCGCTGACGATATCG	46.68	46.68	284	0	320	6	1
66135 TTAGCTCATTCCTGTTACCGGCTGCGGCGCGTACAGCGGATGAAGTGCATCGCG	46.68	46.68	284	0	320	6	1
663 CGTTAGGGGAAGCGTTGTTAGCCAGTTCAGAAAACCGCATTAACATCATATGTCGTG	46.68	46.68	284	0	320	6	1
66195 AAGCAGGTAATCGTGTGCTGGCGGTGAGAAAAGATGCGCATGAATGAATGTCGACTG	46.68	46.68	284	0	320	6	1
743 CAAGATACGACGACGCTTTGCAAGACGAGACAGCTGCTAAAGTGGCGGCGATCCGCA	46.68	46.68	284	0	320	6	1
66255 AAGCATGAAAGAGTACTGCGCGGAAGCAGTAGTAGATTACACGTTCTTCTTCCAC	46.68	46.68	284	0	320	6	1
803 GTTTACTCAAAAATTAAGCAAGTTTCAGATTGTTACACACCAATTAACAGGAGCATTTGCG	46.68	46.68	284	0	320	6	1
66315 AGCTGATCACCGCGGACGCGTGTGCGATCTGCAATCTGCTTTGAAGTAAGCAAT	46.68	46.68	284	0	320	6	1
863 CACATTTAAGTGTACCGCGGATTTGTTGACCGCTTGTCATTCACACACGACACTGGTGGCG	46.68	46.68	284	0	320	6	1
66375 CGCAAGAAAGACGACTGACTGTGCGCTGTCTGCTCATCCGACCCCGCGTGAACGGTT	46.68	46.68	284	0	320	6	1
923 TCCCAAGTGAAGCGCGCTGTATTCATTTGGGACCCATGAGAGAACACACTGTGGCTGTG	46.68	46.68	284	0	320	6	1
66435 TCCCGCATCAGGCGCGGACCCAGGTTATTGTTGCAACTGGAACCGTTTCAGCCGCACTGT	46.68	46.68	284	0	320	6	1
983 TTGCAAGTCTATTGGCTATTATTACCGCAGATAAATAGTGGGAATTTGTTGGTCAATCG	46.68	46.68	284	0	320	6	1
66495 TTGGCGGATTTGGCTTTGGTGTGACACCGAAGTAAGCGGCAATGGGTGGTGAACATCG	46.68	46.68	284	0	320	6	1
1043 GTTCCATGTATGTGAATCAAAAGCAGCAGACGAACTTTATTTTGGCTGGTGGGAGATTG	46.68	46.68	284	0	320	6	1

Db 66555 GCTGGCGAAGGTGGCGGAAATCAGTGGCG-----TCTGTTTGGCGGAGCGGGGATTC 66608

QY 1103 TGGCTGACTCCGATGGCAACAAGATATGAGAAGAACTGGGTTGAATTTGAACCATGC 1162

Db 66609 TGGCTGCGTCGTCACCGCTTGGGTGAGTGGCGGAAACAGCGCCCAACTTTCATCATGT 66668

QY 1163 GGCATATGTT 1172

Db 66669 TGAACGTTTT 66678

RESULT 11

AE004272/c

LOCUS AE004272 14027 bp DNA BCT 31-JUL-2000

DEFINITION Vibrio cholerae chromosome I, section 180 of 251 of the complete chromosome.

ACCESSION AE004272 AE003852

VERSION AE004272.1 GI:9656504

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

FEATURES

source

1. 14027

/organism="Vibrio cholerae"

/strain="N16961"

/serotype="O1"

/db\_xref="taxon:666"

/chromosome="I"

/note="biotype: El Tor"

128. 718

/gene="VC1965"

128. 718

/gene="VC1965"

/note="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAFP95113.1"

/db\_xref="GI:9656505"

/translation="MARNDHTRREELVALTDRVKOFLDTHSYHELISLRKAVTMIGYV  
PSTLVNFGNMYLILHVAQTLDLADQAOAVKSTYPKDALYOLAVCYHDFAKRN  
PYRMOLIEHNNNGEILPEMOAQRINNTSMLEDILKVIAPLRSEQEVLAQASRVLMSC  
VHGITLSVDKDFANPEPIDGNALINNLNHYLAW"

722. 2602

/gene="VC1966"

722. 2602

/gene="VC1966"

/note="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAFP95114.1"

/db\_xref="GI:9656506"

/translation="MNKANPTTALLTERRLPYFTTQGLGAFNDNIFKNVLLFVAR  
AGPQALPISSNLEIFNLAAGLFTLPFLESPAVAGVALLPOOLKESELYRWAKLEIIMLC  
GAIQITQSYGISLEFLPLMGQSAFPEFPVAVALLPOOLKESELYRWAKLEIIMLC  
ILICQITQSYGISLEFLPLMGQSAFPEFPVAVALLPOOLKESELYRWAKLEIIMLC  
TKQITRIKRDRAITQSTMAISWFFELAGVITQTPNNTKILHNGNEAASFLLAFS  
VGIAVSLACDKLSGRIEVLPLGVSIGISFEGYLMATSIIPADLPTEADPVTYS  
ALMPLEAFLLIGFVGVEIVPLVYMLQOQRAKVERAQAIVNALNLYNLFVWGSALG  
IVCLSELEMSIFLEFALLAVNLVAVATIFLOVPFEAVFLWILTHLYVRHNLH  
HIPEQGLIVCNHYSYMDALLISAVCPRLIHPVDEEYANPLIKRLIKAGVPIYS  
ANRRASLRRAPADIPEHSMEGTCVITPEEGCTADGEPNPNRIGDILTRHSPVPV  
IALKGLMSYFSRYKGRACQGVPRFRFSQLEIEAGMPVAPEDQANSAVHEKVAQLRGD  
WR"

2778. 4832

/gene="VC1967"

2778. 4832

/gene="VC1967"

/note="similar to PID:1255679; identified by sequence  
similarity; putative"

/codon\_start=1

/transl\_table=11

/product="methyl-accepting chemotaxis protein"

/protein\_id="AAFP95115.1"

/db\_xref="GI:9656507"

/translation="WSSGKVDLEDSEFMRVKTKILVFSVVGILVTALGVTITKWS  
SSSGISYVGPAMDNDADAMETTIOIEOMLAVNLIGEDSORVEOILNTAIVDVT  
SSSRITIEAGLISAAOTOOFOSQOOSRNLTITSYNTIETKSKYKATQVLDVFG  
EKLETLGSAVELEIREPNRNTTQSDVMERQADGGMESIGLMLKLYTORLDG  
ODDAQOIKRATQAIQROKQANSEMSGTGRTISGEEMKNSYEVFQQAHSQHM  
MAVTSYRNRYOIVETVYTSALIDFIALEELADSKVBOQAVIILDQMAISSEK  
VILITALLVLLGLVNLQIISPIQRQERTVDISENGDITLVNITTDDERELG  
KSFDFEIKRIONLADIYOSTMLATAVDLSATPKVAEAVNKOFEVNTISATTA  
MTAISQVISCAREISQSVLNDKNAOSTLSVROAQOSVELVAVEOGTETISLK  
NHVTIEEVLVDINGIAOTNLALNAIEARAGSGEFAVVADEVRSLTRQGS  
TNTIOESTTORSSADESVRYNNSMLKGTOTTETSQAESIHQVAIEISRTQMO  
QTSSEITHQBOSVYSIASLSHLQALCSAQDKIOQSEHTISALKRKQEDIALKMSK  
K1"

complement(4885. 5430)

/gene="VC1968"

complement(4885. 5430)

/gene="VC1968"

/note="similar to GB:U00096 SP:P7626 PID:1742340  
PID:1742346 PID:1787704; identified by sequence  
similarity; putative"

/codon\_start=1

/transl\_table=11

/product="transcriptional regulator, HTH\_3 family"

/protein\_id="AAFP95116.1"

/db\_xref="GI:9656508"

/translation="MTDWMFKSQIANOLNLRKSRGLSDNTAQTGVSAMKGOIER  
GESSPITAIKMIASGLASFSAPFANPQOLISRSFPDDIMKITHLYLFPVADTGL  
EIFEITLDHDOQMSPHALGVETIYHVLGIMKVFPEQWHEILQGGHIRFSDQPH  
GYAAVTERAVRONIVAYPRRG"

complement(5442. 5537)

/gene="VC1969"

complement(5442. 5537)

/gene="VC1969"

/note="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAFP95117.1"

/db\_xref="GI:9656509"

/translation="MKNPFIMISFGKVVRNSARSEYSSVRSQA"

5518. 6681

/gene="VC1970"

5518. 6681

/gene="VC1970"

/note="similar to GB:M76990 SP:P07775 PID:141751;  
identified by sequence similarity; putative"

/codon\_start=1

574 TTTATTTCAGCAACACCGGAGCGTTAGTGGCTATGTCAGGTGGTCAGATCGCTACGGCG 633

574 TTTATTTCAGCAACACCGGAGCGTTAGTGGCTATGTCAGGTGGTCAGATCGCTACGGCG 633



Best Local Similarity 47.7%; Pred. No. 1.5e-07;  
Matches 287; Conservative 0; Mismatches 297; Indels 18; Gaps 3;

QY	574	TTTATTTCAGCAACACCGGAAACGGTTAGTGCTATGTGTCAGTGCGTCAATTCCTCTACGGCG	633
Db	793	TTTCTTTGGCTCTTTCACCGGAAACGGTTATGTGCGCGCGCTGACAAACCGCTGCGTACTGAA	852
QY	634	GCGGTGCGCTGGGACAAGCCGGCGGGAGCGATGGCGCTGACGATATCGCTTTAGGGCA	693
Db	853	GCGGCTGGCGGGAACGATAGCAAAATATCTGTATGAATAGCAGCGCCAGCAGTATGAGAG	912
QY	694	GCGTTGTATGCCAGTCACAAAAACCGCGCATGTGAACATCAATATGTGCTGGCAAGTATACG	753
Db	913	TGGCTGATGGCGGATGATATAAAACCAAGCGCGGAACATGCTGATGTGAATATCTGT	972
QY	754	ACACGCTTGCAAGACGTGACGACGTGCGCTAAAGGTGGCGGCGCCATGCAAGTTTACTCAMA	813
Db	973	CACGATTTACAGCGCCGATATCCACAGACGCTGGATGTTTTACCGCCGACGACTGTCTGCGT	1032
QY	814	AATAAGCAAGTTGACGATTTGTATCACACCAATTTACAGGGGACATTTGCGGACATTTAACT	873
Db	1033	AAAGTGCAGCATCTTCGCGCGTGTATCTGTGAGCTTCACTCAACAAAGGATGATGATGC	1092
QY	874	GTGACCGGATTTGTGACCGCTTGATCCACACGACACTGGTGGGCTCCACAGTGA	933
Db	1093	TG-----TTTACATCAGTTGTGAGCCAAAGGACGAGTTCGTACCGGCGCAT	1143
QY	934	GCGGCGCTGTATTACATTGCGACCCATGAGAGACACCTCGTGGCTGTTGTCAGATCCT	993
Db	1144	CTGGGGGCAACGATTATATGCCCGCTACAGCAACCGTTCACCCGAGAAATGTATGCCGCTTCT	1203
QY	994	ATTGGCTATTTTACCGCAGATATATGTGGGGAATTTGTGGTTGGCATCCGTTCCATGTAT	1053
Db	1204	GCGGCTCTCTCTCTCAATTCACAAAGCG-----AATTCTGCGTTTCCCTGCGCTCACG	1256
QY	1054	GTGATTCAAAGCAGCAGCAGGACGACCACTTATTTGTGCGGGAGCCGGGATGTGGCTGACTCC	1113
Db	1257	AAAAATTGACGGCATATGTCTGCGA--TTATATGCTTGGCGCGGGGCAATTTCCGTGGTTCC	1314
QY	1114	GATGGCAACAAGATATGAAAGAACTGGGTTGAAATTTGAAACCCATGCGGCAATTTGTTA	1173
Db	1315	GACCCGAGCAAGATGCGACGGAATTCGACAAACAAGCGCGAGGCGTGGCTACTTATTA	1374
QY	1174	AA 1175	
Db	1375	CA 1376	

RESULT 14  
AF078080 LOCUS AF078080 1973 bp mRNA PLN 23-DEC-1998  
DEFINITION Arabidopsis thaliana isochorismate synthase mRNA, complete cds.  
ACCESSION AF078080  
VERSION AF078080.1 GI:3348076  
KEYWORDS  
SOURCE  
ORGANISM  
thale cress.  
Arabidopsis thaliana  
Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II,  
Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 1973)  
Meng, H., Pullman, G.S. and Peter, G.F.  
Cloning of a Plant Isochorismate Synthase (Accession No. AF078080)  
(PCR98-214)  
Plant Physiol. 118 (4), 1536 (1998)  
2 (bases 1 to 1973)  
Meng, H., Peter, G. and Pullman, G.  
Direct Submission  
Submitted (14-JUL-1998) Forest Biology, Institute of Paper Science  
and Technology, 500 10th St, Atlanta, GA 30318, USA  
Location/Qualifiers  
1. 1973  
/organism="Arabidopsis thaliana"

CDS

```

CDS
      /db_xref="taxon:3702"
      217..1728
      /function="catalyzes chorismic acid to isochorismic,acid"
      /note="1csl"
      /codon_start=1
      /product="isochorismate synthase"
      /protein_id="AAC97926.1"
      /db_xref="GI:3348077"
      /translation="MTAVLSPAAATERRLSAVSELKSQPSFSSGVAVLQVPIIDQITG
      AIDMLQANQNETQPCFSPRRSDYGRPDLLDLANDNGNNGNCTYSRRNLVSVAGITQ
      SAVFERIDLPISHDQMSIRKRFSLSTPLRAIYGMRDVGKTAEMERPGATFVS
      PQVEFNERGSSMLAAITAMDELSTWLENIETALQETMLQVGLVLAEMRLRSRGS
      LSKAHVPTKGAVFAVEKALIMINOKSSPLKKVVLNRSARIITPTDIDPIAMLQLOLQ
      EGHAVYFCQLQPGCAPAFIGNTPERLFQRTGLGCESELAATPRPAASSADMDIERI
      LSKPDKDLSEFISREINREKLINCDYVVKPQKTYLAVOHVLSQLAGRLTEKE
      DEYTLIAAHPTPAVCGPGLAEARLLIKETISPGWVACGIFGFGGSEFPAVGRS
      ALVKEGICGALINACTGIVAGSDPSSENNEDLKTISQFKSIEYEATISLQAIN"
BASE COUNT      530 a      392 c      460 g      584 t      7 others
ORIGIN

```

Query Match	5.0%	Score 59.8	DB 7	Length 1973
Best Local Similarity	45.8%	Pred. No. 3.3e-07		
Matches 205; Conservative	0	Mismatches 243	Indels 0	Gaps 0

Oy	574	TTATTTCAGCAACACCGGAAGCGTTAGCGCTATGTACAGCGTGCATGCTCCCTACGGC	633
Db	1096	TTTATTCGGAACACCCCTGAAAGACTATTCCAAAGACTCGATTYAGTGTCTGCAGTGAA	1155
Oy	634	GCAGTCGCTGGGACAAAGCCGGCGGGAGCGATGCGCTGACGATATGCGCTTAGCGGA	693
Db	1156	GCTTTGGCTGCACATAGAGCCCTAGACCTCCTCTTACTGCTCGGATATGAGATAGAGCGT	1215
Oy	694	GCGTTGTTAGGCAGTCAGCAAAAACCGCAATTGCAATCAATATGTCGTGGCAATATACG	753
Db	1216	GACTTACTACACAGTCGGAAGACCACTCGAGTTCTCTATGTACGAGCAATATTAAGA	1275
Oy	754	ACACGCTTGCAAGACGTCAGCAGTCGCTAAAGGTGCCGGCCATCCAAATTTACTCAA	813
Db	1276	GAAAGTTTAAACNGTATATATGTACAGAGTGTGTTGCACAGCCCTCAAAAACCTGTAGGAAG	1335
Oy	814	AATTAAGCAAGTTACAGCATTTTGTACACACCACATTTACAGGGGACATTGGCGGACATTTAAGT	873
Db	1336	CTTGCAAGAGTGCAACATCTATATTTCTCAATTTGGGAGGAGACCTTACGAAGGAATGAT	1395
Oy	874	GTCACCGGATTTGTGATCCGCTTGTCATCCAAACACGAGCACTGGTGGGCTCCACAGTGAA	933
Db	1396	GAGTATTAATAATTTGGCTGCTCTGATCTCCAACTCCAGCTGTTTGTGGGCTTCCACACGA	1455
Oy	934	GCAGCCCTGTATTACATTGCGACCCATGAGAAGACACCTCGTGCGTTGTTGCAGGTCT	993
Db	1456	GAAAGCAAGGCTTTTGTATTAAGAGATTAATCATTCGATAGAGGAATGATGCGGAGCT	1515
Oy	994	ATTGGCTATTTTAAACCGCATATATGTG	1021
Db	1516	ATTGGATTTTGTGCGGAGAGAGTG	1543

RESULT	15
LOCUS	CRO6065
DEFINITION	Cro6065 2078 bp mRNA
ACCESSION	AJ006065
VERSION	AJ006065.1 GI:4127687
KEYWORDS	isochorismate synthase.
SOURCE	Madaagascar periwinkle.
ORGANISM	Catharanthus roseus
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Gentianales; Apocynaceae; Catharanthus.
REFERENCE	1 (bases 1 to 2078)
AUTHORS	van Tegelén, L.J., Moreno, P.R., Croes, A.F., Verpoorte, R. and

TITLE Willems, G. J.  
Purification and cDNA cloning of isochlorismate synthase from  
elicited cell cultures of *Catharanthus roseus*  
JOURNAL Plant Physiol. 119 (2), 705-712 (1999)  
MEDLINE 99137924  
REFERENCE 2 (bases 1 to 2078)  
AUTHORS van Tegelen, L.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAY-1998) van Tegelen L., Department of Experimental  
Botany, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen,  
THE NETHERLANDS

## FEATURES

source

Location/Qualifiers

1..2078

/organism="Catharanthus roseus"

/strain="G. Don (Madagascar periwinkle)"

/db\_xref="taxon:4058"

31..1773

/EC\_number="5.4.99.6"

/codon\_start=1

/product="isochlorismate synthase"

/protein\_id="CA06837.1"

/db\_xref="GI:412768"

/db\_xref="SPTREMBL:Q92PC0"

/translation="MASITGHCVAFHTDLSRKSSEFSNSNNSLFRKSTNIYTRK  
KYRCSTSLSMNGCMGDRAPVGTETETRTLPAYSPALMERLSAVANLSTLPSAQ  
SGIIRLEVPFIEBHIALDMLHSQDKNLIPRCYRSGQVTFSDTSLNLRNCSAA  
NGHLQRISTSSDDKNLVAVAGSVLFRSPNPFESDMLSKRFLSKNCPILRAYGA  
IRFDPARPIAPEWKAQSFYFVPOVEDELHGSMAIATVAMDNALSLTYQOALVAL  
QTTMEQVSVTSKLRQDVSHTSIVSKANIPTRTSMDLTINRYLEEIGNKYSPLTKVL  
ARRSOVITTSIDIDPLAMLSFRADKDAYOCLQHPAFAFIGNTEQLEGRDOLTF  
SEALAAATRARESDSLDLOMAHDLFSSPSLNHEFAIVRENIROKLDATCTSVETPMK  
SVRKIKRIQHLIYARFARGLRSEDEDFKLLSLHPTIYAVCGFPMEDARKFIENEMFDR  
GLIVGVGFEGASQDSFVSIGRSALIGKDALIYAGLVVEGSDPALEMDLELKAS  
QFMKLMKLEAPALK"

BASE COUNT 595 a 423 c 418 g 642 t  
ORIGIN

Query Match 4.8%; Score 57; DB 8; Length 2078;  
Best Local Similarity 50.1%; Pred. No. 2.5e-06;

Matches 172; Conservative 0; Mismatches 165; Indels 6; Gaps 1;

```
OY 806 TACTCAAAATAATGACAGTTGACAGTTGTACACACCAATTACAGGAGACATTGCGGCAC 865
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1385 TAAGAAAGCTTAAGAAATTAACAATCTTATGCTCGATTTGCAGGACAGATTACGCTCG 1444

OY 866 ATTTAGTGTGACCGGATTTGACCGCTTGATCCACACGACGACTGGTGCGCTCC 925
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1445 AAGATGATGAGTTCAAGATTGTTGCTTCCTTCATCTCAGCTGTTTGTGGGTTTC 1504

OY 926 CACGTGAAGCGCCCTGTATTACATTGCGACCCATGAGAACACCTCGTGCCTGTTTG 985
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1505 CTATGGAAGATCGACGGAATTTATTGCGGAAATGAATGTTTGACGAGATTATACG 1564

OY 986 CAGGTCCATTGGCTATTTTACCGCAGATAATAGTGGGAATTTGCTGGATCCGTT 1045
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1565 CTGGCCCTGTTGTTTGGAGAGCTCAGAGTATTTTCTGTTGGAATTAAGATCTG 1624

OY 1046 CCATGTATGTGAATCAAGCAGCAGCACTTATTGCTGTCGCGGATGTGG 1105
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1625 CC-----TTGATTTGGAAGAGATGCCGGTGCTATTAATATATGCGGCGCTTGGGTTGTAG 1678

OY 1106 CTGACTCCGATGCGCAACAAGATATGAAGAAACTGGGTTGAA 1148
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1679 AAGGAAGTATCCAGCTCTAGAAATGCGAGAACTAGAGCTCAA 1721
```

Search completed: November 25, 2000, 01:55:45  
Job time: 2265 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2000, 01:03:28 ; Search time 94.96 Seconds  
(without alignments)  
4735.344 Million cell updates/sec

Title: US-08-774-104A-1

Perfect score: 1197  
Sequence: 1 ATGATGACATATCAGCAAC.....ACTACATCATGTCGATGA 1197

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq.36:.\*  
1: /SID56/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1197	100.0	1197	19	VA1196	Human neutral sph
c	130.6	10.9	4328	18	V41590	Staphylococcus aur
3	102.2	8.5	11764	20	X12985	Enterococcus faeca
4	76	6.3	1659	20	Z23756	E. coli entC DNA.
5	57	4.8	2078	20	Z23758	C. roseus ICS DNA.
c	42.2	3.5	10357	19	V52324	Streptococcus pneu
7	41.6	3.5	5057	20	Z23757	Streptococcus pneu
8	36.2	3.0	1664976	19	V21209	P. fluorescens DNA
c	33.8	2.8	1167	19	Z96384	Mechanococcus jann
10	32.4	2.7	3647	18	V74332	S. pneumoniae derf
11	31.6	2.6	16870	20	X13035	Staphylococcus aur
12	30.8	2.6	2032	13	Q29937	Enterococcus faeca
						PEP4 gene encoding

13	30.8	2.6	3640	20	X88786	Cyclorella cryptic	
14	30.6	2.6	2161	19	Z30444	Tobacco anthranila	
c	15	30.4	2.5	13926	19	V52138	Streptococcus pneu
16	30.2	2.5	231	18	T68178	H. pylori cell env	
17	30.2	2.5	648	19	X30483	H. pylori cellular	
c	18	30.2	2.5	7986	20	Z23937	T. versicolor lacc
19	29.8	2.5	1275	20	Z12106	Neisseria meningit	
20	29.8	2.5	1575	20	Z12105	Neisseria meningit	
c	21	29.6	2.5	1478	20	X07180	Corn threonine dea
22	29.6	2.5	21126	20	X12960	Enterococcus faeca	
23	29.6	2.5	21170	20	X20535	Polynucleotide seq	
c	24	29.4	2.5	1383	9	N80872	Nucleotide sequenc
25	29.4	2.5	1419	16	O89692	pepr. Lactococcus	
26	29.4	2.5	1650	20	Z30446	Tobacco anthranila	
27	29.4	2.5	2380	19	V68834	Zea mays soluble s	
28	29.4	2.5	11597	20	X13159	Enterococcus faeca	
c	29	29.4	2.5	13085	20	X13316	Enterococcus faeca
30	29.4	2.5	24417	18	T97221	Pseudomonas aerugi	
c	31	29.2	2.4	540	21	A05526	Streptococcus pneu
32	29.2	2.4	554	21	Z37314	Human ion transpor	
c	33	29.2	2.4	701	20	X60243	Domain IV sequence
34	29.2	2.4	2573	18	T77804	CDNA encoding vari	
c	35	29.2	2.4	3023	18	T67241	Pasteurella haemol
c	36	29.2	2.4	3336	21	Z46368	Haemorrhagic enter
c	37	29.2	2.4	4503	20	X13086	Enterococcus faeca
38	29.2	2.4	6236	19	X75759	Human neurofilamen	
c	39	29.2	2.4	6524	18	T77803	CDNA encoding wild
c	40	29.2	2.4	6527	18	T77806	CDNA encoding vari
c	41	29.2	2.4	7052	18	T77805	CDNA encoding vari
c	42	29.2	2.4	10254	19	V52162	Streptococcus pneu
c	43	29.2	2.4	26270	21	Z46385	Haemorrhagic enter
44	29	2.4	247	19	V27388	Streptococcus pneu	
45	29	2.4	1119	12	Q12159	Truncated Met-stre	

## ALIGNMENTS

RESULT	1	
VA1196	VA1196 standard; CDNA; 1197 BP.	
XX	XX	
AC	VA1196;	
XX	XX	
DT	20-OCT-1998 (first entry)	
XX	XX	
DE	Human neutral sphingomyelinase encoding CDNA.	
XX	XX	
KW	Human; neutral sphingomyelinase; N-SMase; diagnosis; Crohn's disease;	
KW	obesity; diabetes; Alzheimer's disease; ss.	
XX	XX	
OS	Homo sapiens.	
XX	XX	
FH	Key	Location/Qualifiers
FT	CDS	1..1197
FT		/*tag= a
FT		/product= "neutral sphingomyelinase"
FT		/transl_except= (pos:175..180,aa:Val)
FT		/note= "appears to encode 2 Val residues but only one"
FT		is given in the amino acid sequence (W69162)."
XX	XX	
PN	W09828445-Al.	
XX	XX	
PD	02-JUL-1998.	
XX	XX	
PF	23-DEC-1997;	97MO-US24051.
XX	XX	
PR	24-DEC-1996;	96US-0774104.
XX	XX	
PA	(UYJO ) UNIV JOHNS HOPKINS.	
XX	XX	
PI	Chatterjee S;	
XX	XX	



[illegible]

Db	2582	AAAAATGATTCATTATTATCACTTGTACACTGTAAATAAAGGGCCACCTGAAGATGATTGC	2523
Qy	871	AGTGTGACCGCGGATTTGTAGCCGCTTGATCCAAACACGACATGGGATGCCATCGT	930
Db	2522	TATATTAGTTAAATATGATAA---TTTACATCCAAACACCTCGCTTAGTGGCTATGCCAAA	2466
Qy	931	GAAGCGGCCCTGTATTACATTGCGACCATGAGAAACACTCGTGGCTTGTTCAGAGT	990
Db	2465	GAATTTGGAAAGGATTTTATTATGACGAGAAAGAAATTTGGTAGCGGAGATTTATATGTCGA	2406
Qy	991	CGTATTTGGCTTTTATTAACGCGAGATTAATATGATGGGCAATTTGGTTGGCATCGCTTCATG	1050
Db	2405	CCGGTTGGCTTATANTAGATATATATACGATATATGTGATTTATTTGTGCATTCGTTCATG	2346
Qy	1051	TATGTGAATCCAAACGACGACGAGCAACTTATTTGCTGGTCCGGATTTGTGGCTGAC	1110
Db	2345	CTTATTATGAAGAAGC-----ACAAGCAACTTATTTTGTGGGTGTGGATTTGTAAGAT	2292
Qy	1111	TCCGATGCGGCACAAAGAAATATGAAAGAACTGGGTTGAAATTTGAATCCCATCGGCAATTG	1170
Db	2291	TCTGTATCCAGATATGAGATTTGGCAGAAACGAACCTTAGTTCACACCTATGATGATGATCA	2232
Qy	1171	TTAAGGACTACAAATCATGTGTGAATGA	1197
Db	2231	TTAGGATGTGATATGAAATGGAGAAATCA	2205
RESULT 3			
ID	X12985		
XX	X12985	standard; DNA; 11764 BP.	
AC	X12985;		
XX			
DT	19-MAR-1999	(first entry)	
XX			
DE	Enterococcus faecalis genome contig SEQ ID NO:48.		
XX			
KW	Enterococcus faecalis; contig; detection; Enterococcal infection;		
KW	vaccine; attenuation; computer readable medium; ds.		
OS	Enterococcus faecalis.		
XX			
XX	W09850555-A2.		
XX			
PD	12-NOV-1998.		
XX			
PE	04-MAY-1998; 98WO-US08985.		
XX			
PR	14-NOV-1997; 97US-0066009.		
PR	06-MAY-1997; 97US-0044031.		
PR	16-MAY-1997; 97US-0046655.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Barash SC, Dillon PJ, Kunsch CA;		
XX			
DR	WPI; 1999-045171/04.		
XX			
PT	New isolated Enterococcus faecalis polynucleotides and polypeptides		
PT	used to develop products for the detection of Enterococcus and for		
PT	use in vaccines for prevention or attenuation of Enterococcus		
PT	infection.		
XX			
PS	Claim 1; Page 438-444; 2084pp; English.		
XX			
CC	A computer readable medium has been developed which has recorded on it		
CC	982 nucleotide sequences isolated from the Enterococcus faecalis genome		
CC	X12938 to X13919 represent these nucleotide sequences which are primary		
CC	nucleotide sequences, also known as contigs. The computer-based system		
CC	can identify fragments of the Enterococcus faecalis genome with		
CC	commercial importance. The products can be used to detect the presence		
CC	of Enterococcus faecalis in samples. They can also be used for		
CC	diagnosing Enterococcal infection in an animal and monitoring		

progression of disease, and for identifying agents which can be used to modulate the growth or pathogenicity of *Enterococcus faecalis*, or another related organism, in vivo or in vitro. In particular the CC polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences can be used in vaccines to prevent or attenuate an Enterococcal infection.

Sequence 11764 BP; 3614 A; 2155 C; 2554 G; 3428 T; 13 other;

Query Match 8.5%; Score 102.2; DB 20; Length 11764;  
Best Local Similarity 49.0%; Pred. No. 2,6e-23;  
Matches 365; Conservative 0; Mismatches 368; Indels 12; Gaps 3;

```

423 CTTAGCATGATCAATCAACCTTACCAAGTCGTTTGGTGGGACACAGACCCCTGCAGTT 482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4727 cttagaagaagatgccccttaaaaaagttgttgcgtcgacaactcgtcgtgagcag 4786
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
483 ATCCGACACGTTAGCACTGGCACAATTTATGTCGGTTA--GCTGAGCAGGCAATAC 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4787 tccctgaanaaatcgaagtaaccagtgctcgttaatttaigtgacaacaanaaatac 4846
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
540 GATATCATGTGTTTAAACGACATGATGATTTATTATTCAGCAACACCGAGGTT 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4847 gtaattgttctttagaagccaagaacaagttcattcattgtgcacaccagaacgact 4906
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
600 AGTGGCTATGAGGTGTGCATGCTACGGGCGGCGGTGCTGGGACAAAGCCGGCGG 659
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4907 tctgttgtaactaaagaacgtttgcgacgctgtatgtcgcagaaacattaaacggg 4966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
660 GACGATGGCGCTGACGATATGCGCTTAGGCGCAAGCGTTTATAGCCATGCAAAAAACG 719
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4967 tcaaacaccagaagaagaacaaacttagagagacaactttacaagaccgcaaaaatac 5026
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
720 CATGAACTCATATGTCGTGGCAAGTATACGACACGCTTCAGACAGCTGACGACGTC 779
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5027 cggagaacaccacaattgltgtaacglttagcgaagaagltgycgaagatgcagactc 5086
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
780 GGTAAAGTGGCGGCGCATGCAAGTTTACTGCAAAATTAAGCAAGTTGACATTTGTACAC 839
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5087 tgaanaacaglatccaagcaaccattatcttggaataatcgagatgtaacaacattatgt 5146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
840 ACCAATTACAGGGGACATTTGCGGCACATTTTAACTGTACCGCGATTTGTTGACCGTTGCA 899
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5147 gccaatcgggttcaacgtaaacgggaatttcatttttgaaagtgatgtagcgtaca 5206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
900 TCCACACCAACGACACTGGGTGGCGCTCCACGTCGAAGCGCCCTTATTACTTCGACCCA 959
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5207 cccaacgccaagctttagtgggtgagcccaagaaggttcgcagtcgaatgtaattcgacaata 5266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
960 TGAAGAAGACACGTCGTTGTTTGCAGTCTTATTTGCTATTTTACCG--CAGATTA 1016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5267 tgaagcagaagagtcgtgggtttatattggtgcgaatcogtltgatttcagggaaatgta 5326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1017 TAGTGGGAATTTGTTGTCGATCCGTTCCATGATGTGAATCAAAACGACGACGAGC 1076
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5327 tagtggcgaggttcgctggtgccttgctcgaaggtgtccttgcgtgcagca-----aag 5380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1077 AACTTATTGCTGTCGCGCGGATTTGGCTGACTCCGATGCCACACAGATATATGAGA 1136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5381 cgttctttagcaggttggtgttctgttcgcgaattcccaagcagagctagaagaaga 5440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1137 AACTGGTGAATTTGAACCCATG 1161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5441 aaacgaataaataattcaacgagt 5465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 4  
223756  
ID 223756 standard; DNA; 1659 BP.  
XX  
AC 223756;  
XX

14-JAN-2000 (first entry)  
E. coli entc DNA.  
Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;  
pathogen inducible promoter; antipathogenic protein; toxin;  
antifungal protein; albumin-type protein; hypersensitive response;  
entc; ss.

Escherichia coli.

Key	Location/Qualifiers
CDS	380..1535
FT	/tag= a
FT	/product= "entc"

MO950423-A2.

07-OCT-1999.

25-MAR-1999; 99MO-EP02176.

31-MAR-1998; 98US-0080203.

03-APR-1998; 98US-0080625.

(MOGE-) MOGEN INT NV.

(UYLE-) RIJSDUNIV LEIDEN.

(UYNT-) UNIV NIJMEGEN.

Linthorst HJM, Verpoorte R, Verberne MC, Moreno PRH,

Van Tegelen LJP, Willems GJ, Croes AF, Stuiver MH, Custers J;

Simons LH, Melchers LS, Bol JF;

WPI: 1999-610856/52.

P-PSDB; Y50335, Y33698

Method for inducing pathogen resistance in plants

Claim 5; Page 51-53; 66pp; English.

This invention describes a novel method for the induction of pathogen resistance in plants, by transformation with an expression cassette harboring a gene coding for an isochorismate synthase (ICS). A pathogen inducible promoter can be used to drive expression of a heterologous protein. The heterologous protein used in the method of the invention is an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins, saccharide oxidase, oxalate oxidase, magainins, toxins from *Bacillus thuringiensis*, or antifungal proteins isolated from *Mirabilis jalapa*, *Amaranthus*, *Raphanus*, *Brassica*, *Sinapis*, *Arabidopsis*, *Dahlia*, *Cnicus*, *Lathyrus*, *Clitoria*, *Allium* seeds, *Aralia* and *Impatiens* and albumin-type CC gliadin and wheat-alpha-amylase, or a protein that can induce a hypersensitive response, such as Cf, Bs3 and Po proteins from tomato and CC N-protein from tobacco. This sequence encodes the *Escherichia coli* isochorismate synthase entc gene which is described in the method of the invention.

Sequence 1659 BP; 379 A; 443 C; 454 G; 383 T; 0 other;

Query Match 6.3%; Score 76; DB 20; Length 1659;  
Best Local Similarity 46.6%; Pred. No. 4.8e-15;  
Matches 284; Conservative 0; Mismatches 320; Indels 6; Gaps 1;

```

563 ATGATGAATGTTTATTTCAGCAACACCGGATTTAGTGGGTATGTCAAGTGTGAG 622
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
939 atggtgcgtccctgcgtcgggagccagccggaactgcgtcgaagaaagacggaagcgtc 998
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
623 TCCGTACGGCGGCGGTGTCGGACAAGCCGGCGGAGGATGCGCTGACGATATTCG 682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
999 ttagctccattcgttagccggttcgcgctgcgcagccggaatgaagatgctcgtatcgcg 1058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
683 CGTTAGCGAAGCGTTGTTAGCCATGCAAAAAACCGCATTTGAACATCATATGTCGTGG 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
Db 1059 aagcaggtatcgtctgctggcgctcagaaaaagatcgcataactgtaactggtgtaacc 1118
Qy 743 CAAGATATCGACGACCGCTTGCAAGACGTGACGAGTGGCTAAAGTCCGGCCATCCCA 802
Db 1119 aggcagtaagaagagtagctgcgcaacgagtagtgaattacacgttctcttccac 1178
Qy 803 GTTTACTCAAAATAGCAAGTTACAGTATTGTACACCAATATACAGGGACATTCGG 862
Db 1179 agctgataccacgcgcagcgtctgtgcatctgcgaactcccttgaagtaagcgaatt 1238
Qy 863 CACATTATAGTGTGACCGCGATTTGTGACCGCTTGCAATCCCAACAGACAGATGGTGCG 922
Db 1239 cgcagaagaacgcagcagctctgctgtctgtcatccgaccccgcgctgagcggtt 1298
Qy 923 TCCACAGTGAAGCGCCCTGATTACATTGCGACCCATGAGAAGACACCTCGTGCGTTG 982
Db 1299 tcccgatcagcgcgagccaggtattctgtgaactggaacccgttcgaccgcggaactgt 1358
Qy 983 TTGCAAGTCTATTTGCTATTTTACCGCAGATATAGTGGGGAATTTGTGCGATCC 1042
Db 1359 ttggcgagcatctgtggtctgtgcaagcgaagtaacgagcgtggtgtaaccatcc 1418
Qy 1043 GTTCCATGATGTGAATCAACGACGACGACGACCACTTATTTGCGTGGCGGATTG 1102
Db 1419 gctcgcgagaagctgcgagaaatcagtgcg-----tctgttgcgagagcgggatgtg 1472
Qy 1103 TGGCTGACTCCGATCGCAACCAAGATATGAGAACTGGGTTGAATTTGAACCCATGC 1162
Db 1473 tgcctgcgtcgtcaaccgttggtgagtgcgcgaaacagcgccaacttctacatgt 1532
Qy 1163 GGCATTTGTT 1172
Db 1533 tgaacgtlct 1542
```

```
RESULT 5
ID 223758 standard: DNA; 2078 BP.
AC 223758;
XX 14-JAN-2000 (first entry)
DE C. roseus ICS DNA.
XX Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;
KW pathogen inducible promoter; antipathogenic protein; toxin;
KW antifungal protein; albumin-type protein; hypersensitive response;
KW isochorismate pyruvate synthase; ss.
XX Catharanthus roseus.
OS
XX
PH Key Location/Qualifiers
FT CDS 31..1773
FT /tag= a
FT /product= "ICS"
XX
PN WO9950423-A2.
XX 07-OCT-1999.
XX
XX 25-MAR-1999; 99WO-EP02176.
XX
XX 31-MAR-1998; 98US-0080203.
XX 03-APR-1998; 98US-0080625.
XX
XX (MOGE-) MOGEN INT NV.
XX (UYLE-) RIJKSONIV LEIDEN.
XX (UYNI-) UNIV NIJMEGEN.
XX
PI Linhorst HJM, Verpoorte R, Verberne MC, Moreno PRH;
PI Van Tegelen LJP, Willems GJ, Croes AF, Stuijver MH, Custers J;
```

```
PI Simons LH, Melchers LS, Bol JF;
XX WPI; 1999-610856/52.
DR P-PSDB; Y50340.
XX
PT Method for inducing pathogen resistance in plants
XX
XX Claim 5; Page 60-62; 66pp; English.
XX
CC This invention describes a novel method for the induction of pathogen
CC resistance in plants, by transformation with an expression cassette
CC harboring a gene coding for an isochorismate synthase (ICS). A pathogen
CC inducible promoter can be used to drive expression of a heterologous
CC protein. The heterologous protein used in the method of the invention is
CC an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins,
CC saccharide oxidase, oxalate oxidase, magalins, toxins from Bacillus
CC thuringiensis, or antifungal proteins isolated from Mirabilis jalapa,
CC Catharanthus, Raphanus, Brassica, Sinapis, Arabidopsis, Dahlia, Citrus,
CC Lathyrus, Clitoria, Allium seeds, Aralia and Impatiens and albumin-type
CC proteins, such as thionine, napin, barley trypsin inhibitor, cereal
CC gliadin and wheat-alpha-amylase, or a protein that can induce a
CC hypersensitive response, such as Cf, Bs3 and Po proteins from tomato and
CC N-protein from tobacco. This sequence encodes the Catharanthus roseus
CC isochorismate synthase ICS protein which is described in the method of
CC the invention.
XX
SQ Sequence 2078 BP; 596 A; 423 C; 417 G; 642 T; 0 other;
```

```
Query Match 4.8%; Score 57; DB 20; Length 2078;
Best Local Similarity 50.1%; Pred. No. 1.2e-08;
Matches 172; Conservative 0; Mismatches 165; Indels 6; Gaps 1;
```

```
Qy 806 TACTCAAAATAGCAAGTTACAGATTTGTACACCAATATACAGGGAGACATTGCGGCAC 865
Db 1385 taagaagaagcttaagaagaattcaacatcttatgtctgattgcaagcagatcgcgtc 1444
Qy 866 ATTTAAGTGTGACCGGATTTGTGACCGCTTGCAATCCCAACGACGACGTGGGCGTCC 925
Db 1445 aagatgatgatgtcaagatttcttcccttactcactcagcgttctgtgggttctc 1504
Qy 926 CAGGTGAGCGGCCGTGATTTACATTGCGACCCATGAGAAACACTCGTGCTTGTG 985
Db 1505 ctatgaagatgcacggaatttcttcgcggaataaagtgaacgttgaccgaggaattac 1564
Qy 986 CAGTCTTATTTGCTATTATTTACCGCAGATATAGTGGGGAATTTGTGTCGATCCGTT 1045
Db 1565 ctggccctgtgttcttcttgagagagctcagagtgatcttctgttggaataagatcgt 1624
Qy 1046 CCATGTATGTGAATCAACGACGACGACGACCACTTATTTGCTGTCGCGGAGATTGTGG 1105
Db 1625 c-----ttgatggaaagatgcgcggtgcattaatataatcgagggcgttggtgtag 1678
Qy 1106 CTGACTCCGATGGCAACCAAGATATGAGAACTGGGTTGAA 1148
Db 1679 aaggaagtgatccagctctagaatgcagaactagatgctcaa 1721
```

```
RESULT 6
ID V52324/C
XX V52324 standard: DNA; 10357 BP.
XX
XX V52324;
XX
XX 23-OCT-1998 (first entry)
XX
XX Streptococcus pneumoniae genome fragment SEQ ID NO:191.
XX
XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX
XX Streptococcus pneumoniae.
OS
```

PN WO9818931-A2.  
XX 07-MAY-1998.  
PD  
XX 30-OCT-1997; 97MO-US19588.  
PE  
XX 31-OCT-1996; 96US-0029960.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;  
PI Kunsch CA, Rosen CA;  
DR WPI; 1998-272225/24.  
XX  
XX Computer-readable medium with recorded Streptococcus pneumoniae  
PT polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus  
PT pneumoniae  
PS  
XX Claim 1; Page 1151-1157; 1409pp; English.  
XX  
CC The present invention describes a computer readable medium which has  
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
CC on it, or a representative fragment or a sequence at least 95% identical  
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a  
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
CC to 391, identifying members of the library which contain sequences  
CC that hybridize to the target sequence and isolating the nucleic acid  
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
CC from an organism, amplifying nucleic acid molecules whose nucleotide  
CC sequence is homologous to amplification primers derived from the  
CC fragment of the S. pneumoniae genome to prime the amplification and  
CC isolating the amplified sequences. The computer readable medium can be  
CC used in a computer-based system for identifying fragments of the  
CC S. pneumoniae genome of commercial importance, or expression modulating  
CC fragments of the S. pneumoniae genome. Products from the present  
CC invention can be used in diagnosis kits and assays, and pharmaceutical  
CC compositions and vaccines for S. pneumoniae.  
CC  
XX  
XX Sequence 10357 BP; 2979 A; 2388 C; 1821 G; 3169 T; 0 other;  
SQ

Query Match 3.5%; Score 42.2; DB 19; Length 10357;  
Best Local Similarity 47.5%; Pred. No. 0.0027;  
Matches 162; Conservative 0; Mismatches 173; Indels 6; Gaps 1;

QY 821 AAGTTCAGCATTGTACACACCAATTACAGGGGACATTCGGCACATTAACTGTGACCG 880  
DB 6152 ACGCATGATGATTTACACAGCGTACTCAAGGGCGCTTTCACGAAGTCACTGCGCAAG 6093  
QY 881 CGATTTGTGACCGGCTTCCACACACGATGCGGTGCGCCACGTAAGGCGCC 940  
DB 6092 ATGCGTTGAAAGCAACGCTTCTGCTGGAACCTTTCTGAGACACAAAGATTGAGCGA 6033  
QY 941 TGTATTACATTTGCGACCCATGAGAAGACACCTGCTGGCTTTTGCAGCTCCTATTGGCT 1000  
DB 6032 TGAACGATATCTATGACTGGAACGAAACGGAAGGAGTATACGCTGGGCAATCGGCT 5973  
QY 1001 ATTTTACCGCAGATTAATAGTGGCAATTTGTGTTGGCATCCGTTCCATGTATGATC 1060  
DB 5972 ACTGTGTGCGACGCGTGAATGATTTGGCATTCGCCATCGCAACCA-----TGAATC 5919  
QY 1061 AAACGAGCGAGCAACACTTATTGCTGCTGGCGGATGTGGTGATCCGATCGCC 1120  
DB 5918 TCAGAAATCAAGAGCCTATGTGACGCTGGGCGAGGAGATTGTACGACTATCTATCGCC 5859  
QY 1121 AACAGAAATGAGAAACCTGGTTGAATTTGAACCATG 1161

DB 5858 AAAACGAATACCAAGAAACCAATTACAAAGCAAAATCTATG 5818

RESULT 7  
ID 223757 standard; DNA; 5057 BP.  
XX  
AC 223757;  
XX  
DT 14-JAN-2000 (first entry)  
XX  
DE P. fluorescens DNA encoding ICS orfA and orfD.  
XX  
KW Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;  
KW pathogen inducible promoter; antipathogenic protein; toxin;  
KW antifungal protein; albumin-type protein; hypersensitive response;  
KW isochorismate pyruvate synthase; ss.  
XX  
OS Pseudomonas fluorescens.  
XX  
FH Key Location/Qualifiers  
FT CDS 207..1382  
FT /tag= a  
FT /product= "isochorismate synthase orfA"  
FT CDS 4516..4851  
FT /tag= b  
FT /product= "isochorismate pyruvate lyase orfD"  
XX  
PN WO9950423-A2.  
XX  
XX 07-OCT-1999.  
XX  
PE 25-MAR-1999; 99MO-EP02176.  
XX  
XX 31-MAR-1998; 98US-0080203.  
PR 03-APR-1998; 98US-0080625.  
XX  
PA (MOGE-) MOGEN INT NV.  
PA (UYLE-) RIJKSUNIV LEIDEN.  
PA (UYNT-) UNIV NIJMEGEN.  
XX  
PI Linthorst HJM, Verpoorte R, Verberne MC, Moreno PRH;  
PI Van Tegelen LJP, Willems GJ, Croes AF, Stuijver MH, Custers J;  
PI Simons LH, Melchers LS, Bol JF;  
XX  
DR WPI; 1999-610856/52.  
DR P-PSDB; Y50336, Y50337, Y500338, Y50339.  
XX  
PT Method for inducing pathogen resistance in plants  
PS  
XX Claim 5; Page 54-58; 66pp; English.  
XX  
CC This invention describes a novel method for the induction of pathogen  
CC resistance in plants, by transformation with an expression cassette  
CC harboring a gene coding for an isochorismate synthase (ICS), a pathogen  
CC inducible promoter can be used to drive expression of a heterologous  
CC protein. The heterologous protein used in the method of the invention is  
CC an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins,  
CC saccharide oxidase, oxalate oxidase, magalins, toxins from Bacillus  
CC thuringiensis, or antifungal proteins isolated from Mirabilis jalapa,  
CC Amaranthus, Raphanus, Brassica, Sinapis, Arabidopsis, Dahlia, Citrus,  
CC Lathyrus, Clitoria, Allium seeds, Avena and Impatiens and albumin-type  
CC proteins, such as chitinase, napin, barley trypsin inhibitor, cereal  
CC gliadin and wheat-alpha-amylase, or a protein that can induce a  
CC hypersensitive response, such as Cf, Bs3 and Po proteins from tomato and  
CC N-protein from tobacco. This sequence encodes the Pseudomonas fluorescens  
CC isochorismate synthase orfA protein and the isochorismate pyruvate lyase  
CC orfD protein which are described in the method of the invention.  
XX  
SQ Sequence 5057 BP; 1171 A; 1554 C; 1302 G; 1030 T; 0 other;

	Query Match	3.5%	Score 41.6:	DB 20:	Length 5057:
	Best Local Similarity	45.1%:	Pred. No. 0.0029:		
	Matches 241:	Conservative 0:	Mismatches 284:	Indels 9:	Gaps 2:
Oy	635	CGGTGCTGGGCAAGACCGGCGGGAGACGATGAGCGGCTGACGATATACGGCTTAGGGGAAG	694		
Db	832	cgcttgcagggtctcagcccaagcgcatctgacccgaagcgacggcggaatgcagact	891		
Oy	695	CGTTTGTAGCCAGTCAGAAAAACCGCATTTGACATCAATATCTCGTGCAAGTATCCAGA	754		
Db	892	ggttctgacatcgcgaaaaagatacactaagcaacacgggtctgtgacccagagacatcgtca	951		
Oy	755	CACGCTTGCAAGACGTGCAGACGCTGCTTAAGGTGCGGCGCATGCCAAGTTACTCAAA	814		
Db	952	gccaaacttgggaaactgtgcacgacttgaaatgtgcggcaagccccctccatcatga	1011		
Oy	815	ATAACCAAGTTACGATTTGTACACACCAATTAACAGGGGACATTTGGG---CACATTTAA	871		
Db	1012	cgccgcgcgtctgcaaccttcgaccccgcatcgaagtaacgtctgcaagaccggcgtgat	1071		
Oy	872	GTGTGACCGCGATTTTGTTCGCTTGCCATCCAAACACACGACACTGGGTGGCGTCCACGTG	931		
Db	1072	cggccttcgagcttcgtcgcgcgttcgaccccaacacgcgtctgtgycgcttccacgcg	1131		
Oy	932	AAGCGCCCTGTATTACATTGCGACCCCATGACAGACACCTCTGTGGCTTTGTTGCAAGTTC	991		
Db	1132	agcgcgcccgcgcgctgcgtctgccttcgtcgaaccttcgacgcgcgcgtcttcacgcgca	1191		
Oy	992	CTATTGGCTATTTTACCGGAGATTAATAGGGGAATTTGTGGTTGGCATCGTTCATCT	1051		
Db	1192	tgggtgggttgtgtgcgatgcccgaggaatgagcgaatggcgttaacgacttcgtgcgca	1251		
Oy	1052	ATGTGATCAAAACGACGACGACGACGACCACTTTATTTGCTGTGCTCGGGAATTTGGCTGACT	1111		
Db	1252	cggctcagcgcaaaacgaagtcgcg-----ctgttcgcgcgcgcgagcgtatcttgaagcct	1305		
Oy	1112	CCGATGCCCAACAAGATATGAGAAACTGGGTTGAATTTAACCACATCGGCGC	1165		
Db	1306	caagcccgacttcgatatgycagagaagtcacgaaccaacttgcaccatcgtcgc	1359		
RESULT	8				
ID	V21209	standard; DNA: 1664976 BP.			
AC	V21209;				
XX	10-NOV-1998	(first entry)			
DT					
XX					
XX					
DE		Methanococcus jannaschii circular chromosome.			
XX					
KW		Methanococcus jannaschii: methanogenic archaeon; circular chromosome;			
RV		genome; autotrophic; extrachromosomal element; identification; ds.			
XX					
OS		Methanococcus jannaschii.			
XX					
PN		W09807830-A2.			
XX					
PD		26-FEB-1998.			
XX					
PF		22-AUG-1997; 97WO-US14900.			
XX					
PR		22-AUG-1996; 96US-0024428.			
XX					
PA		(GENO-) INST. GENOMIC RES.			
XX		(UNIT) UNIV ILLINOIS FOUND.			
PA		(UYO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.			
XX					
PI		Bult CJ, Smith HO, Venter JC, White OR, Woese CR;			
XX		WPI, 1998-169145/15.			
XX					
PT		Complete genome sequence of methano-genic archaeon, Methanococcus			

[illegible]

AC 296384;  
XX 10-APR-2000 (first entry)  
XX  
DE S. pneumoniae derived DNA from ORF #212.  
XX  
XX Treatment; prevention; disease; diagnosis; gene therapy; screening;  
KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.  
XX  
OS Streptococcus pneumoniae.  
XX  
XX W09806734-A1.  
PN 19-FEB-1998.  
XX  
PD 15-AUG-1997; 97WO-US14436.  
XX  
PE 16-AUG-1996; 96US-0024022.  
XX  
PR (SMIK ) SMITHKLINE BEECHAM CORP.  
XX  
PA Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
PI Scodola RK;  
XX  
DR WPI; 1998-159452/14.  
XX  
DR P-PSDB; Y86058, Y86059, Y86060.  
XX  
XX Streptococcus pneumoniae proteins and related DNA - useful for  
PT screening compounds for antibacterial activity  
XX  
PS Claim 4; Page 242; 640pp; English.  
XX  
CC This invention describes novel isolated Streptococcus pneumoniae  
CC polynucleotides (see 296173-296494) and their encoded proteins (see  
CC Y85792-Y86182). The DNA, vectors and host cells described in the method  
CC of the invention are useful for the recombinant expression of the  
CC polypeptides. The polypeptides are useful for treatment or prevention of  
CC disease, or diagnosis of disease related to expression or activity of  
CC such a polypeptide. They can also be used to screen for compounds which  
CC interact with and inhibit or activate such a polypeptide. The  
CC polypeptides (or DNA encoding them, via gene therapy) are also useful  
CC for inducing an immunological response in a mammal. The antagonists are  
CC useful to inhibit such bacterial polypeptides. The polypeptides are  
CC particularly useful to identify antimicrobial compounds and antibiotics.  
CC They are also useful to determine their role in pathogenesis of  
CC infection, dysfunction and disease.  
XX  
XX Sequence 1167 BP; 319 A; 280 C; 228 G; 340 T; 0 other;  
SQ

Query Match 2.8%; Score 33.8; DB 19; Length 1167;  
Best Local Similarity 62.4%; Pred. No. 0.5;  
Matches 53; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1055 TGAATCAACGCGACGACCACTTATTTGCTGCGCGATGCTGCTGACTCG 1114  
DB 1149 TGAATTCACAAAATCAAAAGACCTATGTGCAAGCGTGGGCGAGTGTCTACGACTCTA 1090  
QY 1115 ATGGCAACAGATATGAGAAAC 1139  
DB 1089 TCGCCCAAAAGCAATACCAAGAAC 1065

RESULT 10  
V74332  
ID V74332 standard; DNA; 3647 BP.  
XX  
AC V74332;  
XX  
XX 16-MAR-1999 (first entry)  
DE Staphylococcus aureus contig SEQ ID #21.  
XX

KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.  
XX  
XX Staphylococcus aureus.  
OS  
XX  
XX Key location/Qualifiers  
FH misc-feature 301..360  
FT /tag-a  
FT /note- "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
FT misc-feature 2101..2160  
FT /tag-b  
FT /note- "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
XX  
XX EP786519-A2.  
XX  
XX 30-JUL-1997.  
XX  
XX 07-JAN-1997; 97EP-0100117.  
XX  
XX 05-JAN-1996; 96US-0009861.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
PI Rosen CA;  
XX  
XX WPI; 1997-374922/35.  
XX  
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
PT stored on computer readable medium and used in the production of  
PT anti-S.aureus vaccines  
XX  
XX Claim 1; Page 277-279; 3271pp; English.  
XX  
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
XX of the invention. The DNA sequences are recorded on a computer readable  
XX medium, preferably selected from a floppy or hard disk, random access  
XX memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
XX the S.aureus DNA sequences allows putative functions to be assigned so  
XX that protein-encoding or regulatory regions of commercial, therapeutic or  
XX industrial importance can be obtained. Specifically, sequences which are  
XX likely to encode antigens have been identified and these polypeptides can  
XX be used in a vaccine composition against S.aureus infection. The  
XX polypeptides can also be used in a kit for the immunodetection of  
XX S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
XX including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
XX skin and surgical wound infections, scalded skin syndrome, toxic shock  
XX syndrome, etc. Organisms transformed with the DNA sequences can be used  
XX for recombinant production of the polypeptides. The new DNA sequences  
XX (and their fragments) are useful as primers or probes for isolating  
XX homologues of any of the S.aureus DNA sequences contained on the  
XX computer readable medium.  
SQ

Sequence 3647 BP; 1315 A; 550 C; 619 G; 1037 T; 126 other;

Query Match 2.7%; Score 32.4; DB 18; Length 3647;  
Best Local Similarity 45.3%; Pred. No. 2.8;  
Matches 163; Conservative 0; Mismatches 191; Indels 15; Gaps 1;

QY 811 AAAATAGCAAGTTGACATTTGTACACACCAATTAAGGGGACATGCGCAATTTA 870  
DB 563 aatatgaacgtltaatgcatgaagtgaaagtcacaaagtaaaatacaataaattca 622  
QY 871 AGTGTACCGGCGATTTGACCGCTTGATCCACACGACGACTGTGGCGTCCACG 930



Db 623 tcgccaatgacaggtattgagcaattattaccacagcagtcagtgccacaaa 682  
QY 931 GAAGGGCCCTGATTACATTGCGACCATGAGAGACACCTCGCTGTTTGCAGCT 990  
Db 683 ttacgtgaatgaaagaatataatgaacaatataccataaaggcggttataagtg 742  
QY 991 CCTATTGCTATTATTACGCGAGATATAGTGGAAATTTGTGTGGCATCGCTGCATG 1050  
Db 743 ggtgttgataataatctgtaactaactagatttgcaatcgaatcgaacgag 802  
QY 1051 TATGTGAATCAACCGACGACGACCACTTTATTGCTGTGCGGAGTTGTGCTGAC 1110  
Db 803 atgataagat-----gagcagatatacaacgtagaagcggtgtgtgtatagat 856  
QY 1111 TCCGATGCGCAACGAATATGAGAACTGGGTTGAAATTTGAACCCATGCGCAATG 1170  
Db 857 tctattcctgaaagaactgaaatgaacgaatgaaagcctaagaagcttattgagtg 916

## RESULT 11

X13035 ID X13035 standard; DNA: 16870 BP.

XX AC X13035;

XX DT 19-MAR-1999 (first entry)

XX DE Enterococcus faecalis genome contig SEQ ID NO:98.

KW Enterococcus faecalis; contig; detection: Enterococcal infection;  
vaccine; attenuation; computer readable medium; ds.

XX OS Enterococcus faecalis.

XX PN W09850555-A2.

XX PD 12-NOV-1998.

XX PF 04-MAY-1998; 98WO-US08985.

XX PR 14-NOV-1997; 97US-0066009.

XX PR 06-MAY-1997; 97US-0044031.

XX PR 16-MAY-1997; 97US-0046655.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Barash SC, Dillon PJ, Kunsch CA;

XX DR WPI: 1999-045171/04.

XX PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
- used to develop products for the detection of Enterococcus and for  
use in vaccines for prevention or attenuation of Enterococcus  
infection.

XX PS Claim 1; Page 634-642; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it  
XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
XX X12938 to X13919 represent these nucleotide sequences which are primary  
XX nucleotide sequences, also known as contigs. The computer-based system  
XX can identify fragments of the Enterococcus faecalis genome with  
XX commercial importance. The products can be used to detect the presence  
XX of Enterococcus faecalis in samples. They can also be used for  
XX diagnosing Enterococcal infection in an animal and monitoring  
XX progression of disease, and for identifying agents which can be used to  
XX modulate the growth or pathogenicity of Enterococcus faecalis, or  
XX another related organism, in vivo or in vitro. In particular the  
XX polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
XX can be used in vaccines to prevent or attenuate an Enterococcal  
XX infection.

SQ Sequence 16870 BP; 5332 A; 3410 C; 2947 G; 5177 T; 4 other;

Query Match 2.6%; Score 31.6; DB 20; Length 16870;

Best Local Similarity 52.2%; Pred. No. 13;

Matches 70; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 461 GTCGCAACAGACCTGCGATTATCCGACGTTNCGACTGGCACAAATATTGCGCT 520  
Db 7311 gtcgtaattgactcctggaaatcgtcccaatcgcaacgataatcaaaagtagtcatt 7370  
QY 521 TAGCTGAGCAGCGCAATACGTATCATGTGTTTAAAGCAGACATGTAATGTTATT 580  
Db 7371 gatltccccaatcctaagaacgtgttttcttaaatcaatgttgatcgcttctc 7430  
QY 581 CAGCAACACCGGAA 594  
Db 7431 aggtccatcgta 7444

## RESULT 12

Q29937 ID Q29937 standard; cDNA: 2032 BP.

XX AC Q29937;

XX DT 11-MAR-1993 (first entry)

XX DE PEP4 gene encoding Pichia pastoris proteinase A.

KW Proteolytic; vacuolar; aspartyl protease; EGF; GRF; IGF-1; ss.

XX OS Pichia pastoris.

XX FH Key Location/Qualifiers

XX FT CDS 239..1468

XX FT CDS /\*tag= a

XX PN W09217595-A.

XX PD 15-OCT-1992.

XX PF 31-MAR-1992; 92MO-US02521.

XX PR 01-APR-1991; 91US-0678916.

XX PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.

XX PI Gleeson MA, Howard BD;

XX DR WPI: 1992-366264/44.

XX PT P-PSDB; R28030.

XX PT New isolated DNA fragment from Pichia comprising gene influencing  
proteolytic activity - used to generate proteolytic deficient  
strains, as hosts for expression of proteolytically sensitive  
proteins

XX PS Claim 7; Page 67-69; 108pp; English.

XX The Pichia pastoris PEP4 gene encoding proteinase A was identified in  
XX a lambda based EMBL3 P. pastoris genomic DNA library by its ability  
XX to hybridise with a radiolabelled fragment of the homologous  
XX Saccharomyces cerevisiae PEP4 gene. The P. pastoris PEP4 gene was  
XX cloned by isolating positive plaques contg. the hybridising  
XX recombinant phage DNA. The PEP4 prod., a vacuolar, aspartyl  
XX protease capable of self-activation, as well as subsequent  
XX activation of additional vacuolar proteases, e.g. carboxypeptidase Y  
XX and proteinase B may be mutated to create a proteolysis-deficient  
XX strain of Pichia which may be used for recombinant expression of  
XX proteolytically sensitive proteins, e.g. EGF, GRF and IGF-1.  
XX See also Q29938-41.







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2000, 20:03:37 ; Search time 85.07 Seconds  
(without alignments)  
2128.083 Million cell updates/sec

Title: US-08-774-104A-1

Perfect score: 1197

Sequence: 1 ATGATGACATATCAGCAAC.....ACTACATCATGTCGATGA 1197

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents,NA:\*

1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/CTUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/1/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1197	100.0	1197	4	US-08-774-104A-1
2	33.8	2.8	7218	1	US-08-232-463-14
3	30.8	2.6	2032	1	US-08-088-633-1
4	30.8	2.6	2032	1	US-08-245-756-1
5	30.8	2.6	2032	2	US-08-441-750-1
6	30.8	2.6	2032	3	US-08-441-751-1
7	30.8	2.6	2032	6	PCR-US92-02521-1
8	30.8	2.6	3640	4	US-08-627-873-6
9	30.6	2.6	2161	4	US-09-001-825-4
10	30.6	2.6	2161	4	US-09-001-826-25
11	30.4	2.5	458	5	US-09-141-000-4
12	29.4	2.5	1419	5	US-08-242-098-39
13	29.4	2.5	1650	4	US-09-001-826-24
14	29.4	2.5	2380	2	US-08-572-951-3
15	29.4	2.5	2447	4	US-08-846-762-1
16	29.2	2.4	554	4	US-09-121-179-2
17	29	2.4	1119	1	US-07-854-596B-30
18	29	2.4	1257	1	US-07-854-596B-25
19	29	2.4	1317	1	US-07-854-596B-18
20	29	2.4	1335	1	US-07-854-596B-14
21	29	2.4	1458	1	US-07-854-596B-42
22	29	2.4	1466	4	US-08-749-902-2
23	29	2.4	1467	1	US-07-854-596B-46
24	29	2.4	1512	1	US-07-854-596B-27
25	29	2.4	2253	1	US-07-854-596B-39
26	29	2.4	2589	1	US-07-854-596B-34

27	28.8	2.4	1908	5	US-09-041-991A-3	Sequence 3, Appl
28	28.6	2.4	1079	2	US-08-454-196-16	Sequence 16, Appl
29	28.6	2.4	1079	5	US-09-064-033-16	Sequence 16, Appl
30	28.6	2.4	3164	1	US-07-723-002C-3	Sequence 3, Appl
31	28.6	2.4	5904	2	US-08-309-512-1	Sequence 1, Appl
32	28.6	2.4	5904	6	PCR-US92-08756A-1	Sequence 1, Appl
33	28.4	2.4	2524	1	US-08-317-522A-1	Sequence 1, Appl
34	28.4	2.4	2524	1	US-08-439-818A-1	Sequence 1, Appl
35	28.4	2.4	2524	3	US-08-751-965-1	Sequence 1, Appl
36	28.4	2.4	2524	3	US-08-738-975-1	Sequence 1, Appl
37	28.4	2.4	2524	4	US-08-728-626-1	Sequence 1, Appl
38	28.4	2.4	2524	5	US-08-808-599A-1	Sequence 1, Appl
39	28.4	2.4	3477	1	US-08-141-324-13	Sequence 13, Appl
40	28.4	2.4	3477	2	US-08-541-902-13	Sequence 13, Appl
41	28.4	2.4	6241	2	US-08-570-311-25	Sequence 25, Appl
42	28.2	2.4	4406	1	US-08-369-043-1	Sequence 1, Appl
43	28.2	2.4	43676	5	US-09-356-952-12	Sequence 12, Appl
44	28	2.3	891	7	5164490-2	Patent No. 5164490
45	27.8	2.3	685	5	US-08-792-832A-47	Sequence 47, Appl

#### ALIGNMENTS

RESULT 1

US-08-774-104A-1

Sequence 1, Application US/08774104A

Patent No. 5919687

GENERAL INFORMATION:

APPLICANT: Chatterjee, Subroto

TITLE OF INVENTION: RECOMBINANT N-SMASES AND NUCLEIC ACIDS

TITLE OF INVENTION: ENCODING SAME

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dikey, Bronstein, Roberts & Cushman, LLP

STREET: 130 Water Street

City: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/774,104A

FILING DATE: 12/24/96

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Corless, Peter F.

REGISTRATION NUMBER: 33,860

REFERENCE/DOCKET NUMBER: 46906

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

TELEX:

INFORMATION FOR SEQ. ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1197 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE:

ORIGINAL SOURCE:

US-08-774-104A-1





Query Match 2.6%; Score 30.8; DB 1; Length 2032;  
Best Local Similarity 51.4%; Pred. No. 2.1;  
Matches 71; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 444 AGCCAAAGTGTGTTTGGTGGCAAGACCCCTGCACTTATCCGACACGTTAGCACTGGC 503  
DB 307 AGCCAAAGTGTGTTTGGTGGCAAGACCCCTGCACTTATCCGACACGTTAGCACTGGC 366  
QY 504 ACAATATTATTCGTGCTAGCTGAGCAGCGCAATACGTATCATGTGTTTAAAGCGACA 563  
DB 367 CAATTTGGGCAAGTATCTCTGCTGGAACATTAATATGTTTCTGTTCACAGCAACA 426  
QY 564 TGATGAATGTTTATTTTC 581  
DB 427 AATGCTTTGTCCAAAGTC 444

RESULT 5  
US-08-441-750-1  
; Sequence 1, Application US/08441750  
; Patent No. 5691166  
; GENERAL INFORMATION:  
; APPLICANT: Gleeson, Martin A  
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHTA PROTEOLYTIC  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/441,750  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/245,756  
; FILING DATE: 16-May-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/088,633  
; FILING DATE: 06-JULY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/678,916  
; FILING DATE: 01-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 9763  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2032 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 239..1468  
; FEATURE:

NAME/KEY: mat\_peptide  
LOCATION: 239..1468  
US-08-441-750-1

Query Match 2.6%; Score 30.8; DB 2; Length 2032;  
Best Local Similarity 51.4%; Pred. No. 2.1;  
Matches 71; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 444 AGCCAAAGTGTGTTTGGTGGCAAGACCCCTGCACTTATCCGACACGTTAGCACTGGC 503  
DB 307 AGCCAAAGTGTGTTTGGTGGCAAGACCCCTGCACTTATCCGACACGTTAGCACTGGC 366  
QY 504 ACAATATTATTCGTGCTAGCTGAGCAGCGCAATACGTATCATGTGTTTAAAGCGACA 563  
DB 367 CAATTTGGGCAAGTATCTCTGCTGGAACATTAATATGTTTCTGTTCACAGCAACA 426  
QY 564 TGATGAATGTTTATTTTC 581  
DB 427 AATGCTTTGTCCAAAGTC 444

RESULT 6  
US-08-441-751-1  
; Sequence 1, Application US/08441751  
; Patent No. 5831053  
; GENERAL INFORMATION:  
; APPLICANT: Gleeson, Martin A  
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHTA PROTEOLYTIC  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/441,751  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/245,756  
; FILING DATE: 16-May-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/088,633  
; FILING DATE: 06-JULY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/678,916  
; FILING DATE: 01-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 9763  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2032 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: CDNA  
; FEATURE:



```

; NAME/KEY: CDS
; LOCATION: 239..1468
; FEATURE:
; NAME/KEY: mat-peptide
; LOCATION: 239..1468
;
US-08-441-751-1

```

Query Match	2.68;	Score 30.8;	DB 3;	Length 2032;
Best Local Similarity	51.48;	Pred. No. 2.1;		
Matches 71; Conservative	0;	Mismatches 67;	Indels 0;	Gaps 0

QY	444	AGCCAAAGTCCTTTTGGTCGCGCAACAGACCTTCAGATTATTCGACACAGCTTACGACTGGC	5033
Db	307	AGCCAAAGTTCATCTCGTAAAGATATACCAAGCATCCAGTCTCAGAAACTTTTAAAGAGGC	3666
QY	504	ACAAATTTATTCGTCGTTAGCTAGTCAGCGGCAATACGTATACATGTTGGTTTAAAGGACAC	5633
Db	367	CAATTTTGGCAGATATGCTCTGCTCTGAGCAATTAATATGTTCTCTGTTCACGACACAC	4266
QY	564	TGATGATTTGTTTATTTTC	581
Db	427	AAATGCTTTGTCCAAGTC	444

RESULT 7  
PCT-US92-02521-1  
; Sequence 1, Application PC/TUS9202521

1 GENERAL INFORMATION:  
2 APPLICANT: Gleeson, Martin A  
3 APPLICANT: Howard, Bradley D  
4 TITLE OF INVENTION: GENES WHICH INFLUENCE PICHA PROTEOLYTIC  
5 TITLE OF INVENTION: ACTIVITY, AND USES THEREFOR  
6 NUMBER OF SEQUENCES: 6  
7 CORRESPONDENCE ADDRESS:  
8 ADDRESSEE: Fitch, Even, Tabin & Flannery  
9 STREET: 135 South LaSalle Street, Suite 900  
10 CITY: Chicago  
11 STATE: Illinois  
12 COUNTRY: U.S.A.  
13 ZIP: 60603  
14 COMPUTER READABLE FORM:  
15 MEDIUM TYPE: Floppy disk  
16 COMPUTER: IBM PC compatible  
17 OPERATING SYSTEM: PC-DOS/MS-DOS  
18 SOFTWARE: PatentIn Release #1.0, Version #1.25  
19 CURRENT APPLICATION NUMBER: PCT/US92/02521  
20 APPLICATION NUMBER: PCT/US92/02521

LOCATION: 239..1468  
PCT-US92-02521-1

Query Match	2.68;	Score 30.8;	DB 6;	Length 2032;
Best Local Similarity	51.48;	Pred. No. 2.1;		
Matches 71; Conservative	0;	Mismatches 67;	Indels 0;	Gaps 0;

[illegible]

RESULT 8  
US-08-627-873-6  
; Sequence 6, Application US/08627873  
Data4: US-08627873-6

```

1  GENERAL INFORMATION:
2  APPLICANT: Jarvis, Eric E.
3  APPLICANT: Roessler, Paul G.
4  TITLE OF INVENTION: ISOLATED GENE ENCODING AN ENZYME WITH
5  TITLE OF INVENTION: UDP-GLUCOSE PYROPHOSPHORYLASE AND
6  TITLE OF INVENTION: PHOSPHOGLUCOMUTASE ACTIVITIES FROM
7  TITLE OF INVENTION: CYCLOTELLA CRYPTICA
8  NUMBER OF SEQUENCES: 9
9  CORRESPONDENCE ADDRESS:
10 ADDRESSSEE: National Renewable Energy Laboratory
11 STREET: 1617 Cole Boulevard
12 CITY: Golden
13 STATE: CO
14 COUNTRY: U.S.A.
15 ZIP: 80401-3393
16
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Floppy disk
19 COMPUTER: IBM PC compatible
20 OPERATING SYSTEM: PC-DOS/MS-DOS
21 SOFTWARE: ASCII(DOS)text (*.*)
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/627, 873
24 FILING DATE: April 3, 1996
25 CLASSIFICATION: 435
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Edna M. O'Connor
28 REGISTRATION NUMBER: 29252
29 REFERENCE/DOCKET NUMBER: 95-57
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: 303/384-7573
32 TELEFAX: 303/384-7499
33 INFORMATION FOR SEQ. ID NO.: 6:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 3640 base pairs
36 TYPE: nucleic acid
37 STRANDEDNESS: double
38 TOPOLOGY: linear
39 MOLECULE TYPE: DNA (genomic)
40 FEATURE:
41 FEATURE: 1.
42 NAME/KEY: CDS (exon)
43 LOCATION: 1..24
44 NAME/KEY: Intron
45 LOCATION: 25..314
46 NAME/KEY: CDS (exon)
47 LOCATION: 315..782
48 NAME/KEY: Intron
49 LOCATION: 783..885

```

NAME/KEY: CDS (exon)  
LOCATION: 886..1402  
NAME/KEY: Intron  
LOCATION: 1403..1478  
NAME/KEY: CDS (exon)  
LOCATION: 1479..3637  
US-08-627-873-6

Query Match 2.6%; Score 30.8; DB 4; Length 3640;  
Best Local Similarity 61.0%; Pred. No. 3;  
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 65 TTGAACAACGATTTGGCGCTTACTTTCGACACCCGCTGATGATCTTACGTTTG 124  
DB 2394 TTGACATACCTACGACGAGCTACCACTGTTCAGCCGCTATGAGATACAGTGTG 2453  
QY 125 GCATTGGCGCAATCCGTACGGC 146  
DB 2454 TTGTTGAGGTGATCCCTCGAC 2475

RESULT 9  
US-09-001-826-4  
Sequence 4, Application US/09001826A  
Patent No. 5965727  
GENERAL INFORMATION:  
APPLICANT: SONG, HEE-SOOK  
APPLICANT: BROTHERTON, JEFFREY E.  
APPLICANT: WIDHOLM, JACK M.  
TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE  
FILE REFERENCE: UI001.C1  
CURRENT APPLICATION NUMBER: US/09/001,826A  
EARLIER FILING DATE: 1997-12-31  
EARLIER APPLICATION NUMBER: 08/937,739; 60/025,140  
EARLIER FILING DATE: 1997-07-25; 1996-07-26  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text  
SEQ ID NO 4  
LENGTH: 2161  
TYPE: DNA (cDNA)  
ORGANISM: Nicotiana tabacum  
US-09-001-826-4

Query Match 2.6%; Score 30.6; DB 4; Length 2161;  
Best Local Similarity 49.7%; Pred. No. 2.5;  
Matches 78; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 572 TGTATTTCACGACACCGAAGGTTAGTGGCTATGTCAGTGATCGTACGG 631  
DB 1144 tttagtgcacgcagccagaaatttgacacgltggaagaagaatgttaac 1203  
QY 632 CGGCGCTGCTGGAGCAACCGCGCGGAGCGATGCGCTGACATATCGGTTAGCGG 691  
DB 1204 gaccactgctggaagaagaagaagaagaacacccatgatgagatgtgttgga 1263  
QY 692 AAGCTTTGACCACTCAGAAAACCGCATTCAGCA 728  
DB 1264 tgcagatgtltaagaagatgagaacaacgcgcagagca 1300

RESULT 10  
US-09-001-826-25  
Sequence 25, Application US/09001826A  
Patent No. 5965727  
GENERAL INFORMATION:  
APPLICANT: SONG, HEE-SOOK  
APPLICANT: BROTHERTON, JEFFREY E.  
APPLICANT: WIDHOLM, JACK M.  
TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE  
FILE REFERENCE: TRANSFORMATION

FILE REFERENCE: UI001.C1  
CURRENT APPLICATION NUMBER: US/09/001,826A  
CURRENT FILING DATE: 1997-12-31  
EARLIER APPLICATION NUMBER: 08/937,739; 60/025,140  
EARLIER FILING DATE: 1997-07-25; 1996-07-26  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text  
SEQ ID NO 25  
LENGTH: 2161  
TYPE: DNA (cDNA)  
ORGANISM: Nicotiana tabacum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: LOCATION: 90..1940  
US-09-001-826-25

Query Match 2.6%; Score 30.6; DB 4; Length 2161;  
Best Local Similarity 49.7%; Pred. No. 2.5;  
Matches 78; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 572 TGTATTTCACGACACCGGAAAGGTTAGTGGCTATGTCAGTGATCGTACGG 631  
DB 1144 tttagtgcacgcagccagaaatttgacacgltggaagaagaatgttaac 1203  
QY 632 CGGCGCTGCTGGAGCAACCGCGCGGAGCGATGCGCTGACATATCGGTTAGCGG 691  
DB 1204 gaccactgctggaagaagaagaagaagaacacccatgatgagatgtgttgga 1263  
QY 692 AAGCTTTGACCACTCAGAAAACCGCATTCAGCA 728  
DB 1264 tgcagatgtltaagaagatgagaacaacgcgcagagca 1300

RESULT 11  
US-09-141-000-4/c  
Sequence 4, Application US/09141000  
Patent No. 6054295  
GENERAL INFORMATION:  
APPLICANT: Chen, Pang  
TITLE OF INVENTION: DNA MOLECULES ENCODING HUMAN NUCLEAR  
FILE REFERENCE: 19999y  
CURRENT APPLICATION NUMBER: US/09/141,000  
CURRENT FILING DATE: 1998-08-26  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 458  
TYPE: PRT  
ORGANISM: Human  
US-09-141-000-4

Query Match 2.5%; Score 30.4; DB 5; Length 458;  
Best Local Similarity 9.1%; Pred. No. 1.1;  
Matches 29; Conservative 95; Mismatches 199; Indels 0; Gaps 0;

QY 664 GATGGCGCTGACGATATCGCTTAGCGCAAGCGTTGTTAGCCATCAGAAAAACCGCAT 723  
DB 458 BMT..K...MDK..BMC..M.NR.D.BTWASA.Y...AK.KMCTYY.H.KD.CT.RH.. 399  
QY 724 GAACATCATATATGCTGGCAAGTATCAGACACGCTTGCAAGAGCTGACGACGTCGCTA 783  
DB 398 T.D.BH..M.BT.BH..DKSHSNT.T.TM.AB...M..MKSRRMB...TNN.H..CT.. 339  
QY 784 AAGTGCGCGCCATGCAAGTTTACTCAAAAATAAGCAAGTACACATTTGATACACCA 843  
DB 338 MS.H.HK.RHHTRB..H..S.SYRBC.....KMTS...SK.HT.S.AS.C..DMTWC.. 279  
QY 844 ATTACAGGAGCATTTGCGGACATTTAGTGTGACCGGATTTGACCGCTTGATCA 903  
DB 278 BB..YHT.HG..AA.TM.HSH..BA.H.KTR.M...TB..DSB.MNR.MMT..B...N.R.S 219

QY 904 ACACGACACTGGGTGGCGCTCCACGCTGAAGCGGCCCTGTATTACATTGGACCATGAG 963  
DB 218 N.YH.YYMR.YCCYBYH.YBC.M.KCBM.GMK.YGT.GSMYYMA..G.NAT.GSR..N 159  
QY 964 AAGACACCTCGTGGCTT 980  
DB 158 C.AYM..TMGT.GSTB 142

RESULT 12  
US-08-242-098-39  
; Sequence 39, Application US/08242098  
; Patent No. 5691185  
; GENERAL INFORMATION:  
; APPLICANT: DICKELY, Francoise  
; APPLICANT: JOHANSEN, Eric  
; APPLICANT: NILSSON, Dan  
; APPLICANT: HANSEN, Bogn  
; APPLICANT: STROMAN, Per  
; TITLE OF INVENTION: Lactic Acid Bacterial Suppressor Mutants  
; TITLE OF INVENTION: and Their Use as Selective Markers and as Means of  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/242,098  
; FILING DATE: 13-MAY-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/133,390  
; FILING DATE: 08-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30307/141/PLVI  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1419 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-242-098-39

Query Match 2.5%; Score 29.4; DB 2; Length 1419;  
Best Local Similarity 51.1%; Pred. No. 5.1;  
Matches 69; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 68 AACACACTGAATTTGGCGCTTACTTTCGACACCGCTGATGACTTTACGTTTGCA 127  
DB 584 ATATCAAGATATATTCATTTCTCAGTAAATATGCTGCAAGTGTCTTACAGCT 643  
QY 128 TTGGCCGAATCGCTACGGCAAAAAGCGCTCAGGCATTACAAGTTCGCGTTGTTGGTG 187  
DB 644 TTTAAGCAGCTCTTGCAAGAAATATGTTCCAGATCAGCAACTGCAGTATTTCAAGTGT 703  
QY 188 CGCAGTCATTTGATG 202

DB 704 CTAAGATTAGAAG 718

RESULT 13  
US-09-001-826-24  
; Sequence 24, Application US/0901826A  
; Patent No. 5965727  
; GENERAL INFORMATION:  
; APPLICANT: SONG, HEE-SOOK  
; APPLICANT: BROTHERTON, JEFFREY E.  
; APPLICANT: WIDHOLM, JACK M.  
; TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE  
; TITLE OF INVENTION: TRANSFORMATION  
; FILE REFERENCE: U1001.C1  
; CURRENT APPLICATION NUMBER: US/09/001,826A  
; EARLIER FILING DATE: 1997-12-31  
; EARLIER APPLICATION NUMBER: 08/937,739; 60/025,140  
; EARLIER FILING DATE: 1997-07-25; 1996-07-26  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text  
; SEQ ID NO 24  
; LENGTH: 1650  
; TYPE: DNA (cDNA)  
; ORGANISM: Nicotiana tabacum  
; US-09-001-826-24

Query Match 2.5%; Score 29.4; DB 4; Length 1650;  
Best Local Similarity 46.4%; Pred. No. 5.6;  
Matches 96; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 522 AGCTGACGAGCGGATATCGTATCATGTGCTTTAAAGCGACATGATGATTGTTATTC 581  
DB 742 agtgtgataaccgaagtcgataatgttgcagctcagaagatgcatltygtagc 801  
QY 582 AGCAACACGGGAACGGTTAGTGGCTATGTCAGTGTGTCAGATGCTACGGCGGCGTGC 641  
DB 802 atcaagtcagaatcttcaccacaagaagcagaacaagatgtaatcgtgccaatgyc 861  
QY 642 TGGGACACAGCGCGCGCGAGCGATGCGCTGACGATATGCGCTTAGCGCAAGCGTTGTT 701  
DB 862 aggaaccagaagaagagaggaagatgtaagttgagataagaagatgaataagaactgct 921  
QY 702 AGCCAGTCAGAAAAACCGCATTTGACA 728  
DB 922 agaaatgaataaagaagatgctgagca 948

RESULT 14  
US-08-572-951-3  
; Sequence 3, Application US/08572951  
; Patent No. 5824790  
; GENERAL INFORMATION:  
; APPLICANT: KEPLING, PETER L.  
; APPLICANT: KNIGHT, MARY E.  
; APPLICANT: GUAN, HANPING  
; TITLE OF INVENTION: MODIFICATION OF STARCH  
; TITLE OF INVENTION: SYNTHESIS IN PLANTS  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN  
; ADDRESSEE: Intellectual Property Group of  
; ADDRESSEE: Pillsbury Madison & Sutro LLP  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

1      CURRENT APPLICATION DATA:
2      APPLICATION NUMBER:   US/08/572,951
3      FILING DATE: 15-DEC-1995
4      CLASSIFICATION: 800
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: 08/346,602
7      FILING DATE: 29-NOV-1994
8      CLASSIFICATION: 800
9      PRIOR APPLICATION DATA:
10     APPLICATION NUMBER: 08/263,921
11     FILING DATE: 21-JUN-1994
12     CLASSIFICATION: 800
13     ATTORNEY/AGENT INFORMATION:
14     NAME: Paul N. Kokulis
15     REGISTRATION NUMBER: 16,773
16     REFERENCE/DOCKET NUMBER: 222957/1..02.15C
17     TELECOMMUNICATION INFORMATION:
18     TELEPHONE: (202) 861-3000
19     TELEFAX: (202) 822-0944
20     INFORMATION FOR SEQ ID NO: 3:
21     SEQUENCE CHARACTERISTICS:
22     LENGTH: 2380 base pairs
23     TYPE: nucleic acid
24     STRANDEDNESS: both
25     TOPOLOGY: linear
26     MOLECULE TYPE: CDNA
27     US-08-572-951-3

```

Query Match	2.58	Score	29.4	DB	2	Length	2380
Best Local Similarity	30.18	Pred	No. 7				
Matches	44	Conservative	27	Mismatches	75	Indels	0
						Gaps	0

[illegible]

```

RESULT 15
US-08-846-762-1
: Sequence 1, Application US/08846762A
: Patent No. 5994072
: GENERAL INFORMATION:
: APPLICANT: Lam, Joseph S.
: APPLICANT: Burrows, Lori
: APPLICANT: Charter, Deborah
: APPLICANT: de Kievit, Teresa
: TITLE OF INVENTION: No. 5994072e1 Proteins Involved in the Synthesis and Assembly
: TITLE OF INVENTION: of O-antigen in Pseudomonas Aeruginosa
: FILE REFERENCE: 6580-089
: CURRENT APPLICATION NUMBER: US/08/846,762A
: CURRENT FILING DATE: 1997-04-30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO. 1
: LENGTH: 24417
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-08-846-762-1

```

Query Match	2.5%	Score 29.4	DB 4	Length 24417
Best Local Similarity	51.1%	Pred. No. 30		
Matches 69	Conservative 0	Mismatches 66	Indels 0	Gaps 0

[illegible]

Search completed: November 25, 2000, 01:48:46  
Job time: 20709 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2000, 18:38:06 ; Search time 852.26 Seconds  
(without alignments) 8683.758 Million cell updates/sec

Title: US-08-774-104A-1

Perfect score: 1197  
Sequence: 1 ATGATGACATATCAGCAAC.....ACTACATCATGTGCAATGA 1197

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estda:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estom:\*  
68: em\_estov1:\*  
69: em\_estov2:\*  
70: em\_estpl1:\*  
71: em\_estpl2:\*  
72: em\_estpl3:\*  
73: em\_estpl4:\*  
74: em\_estpl5:\*  
75: em\_estro1:\*  
76: em\_estro2:\*  
77: em\_estro3:\*  
78: em\_estro4:\*  
79: em\_estro5:\*  
80: em\_estro6:\*  
81: em\_estro7:\*  
82: em\_estro8:\*  
83: em\_estro9:\*  
84: em\_estro10:\*  
85: em\_estro11:\*  
86: em\_estro12:\*  
87: em\_estro13:\*  
88: gb\_gss1:\*  
89: gb\_gss2:\*  
90: gb\_gss3:\*  
91: gb\_gss4:\*  
92: em\_gss1:\*  
93: em\_gss2:\*  
94: em\_gss3:\*  
95: em\_gss4:\*  
96: gb\_gss5:\*  
97: gb\_gss6:\*  
98: gb\_gss7:\*  
99: gb\_gss8:\*  
100: gb\_gss9:\*  
101: em\_gss5:\*  
102: em\_gss6:\*  
103: em\_gss7:\*  
104: em\_gss8:\*  
105: em\_gss9:\*  
106: em\_gss10:\*  
107: em\_gss11:\*  
108: gb\_gss10:\*  
109: gb\_gss11:\*  
110: em\_gss12:\*  
111: gb\_gss12:\*  
112: gb\_gss13:\*  
113: gb\_gss14:\*  
114: gb\_gss15:\*  
115: gb\_gss16:\*  
116: gb\_gss17:\*

117: gb-gss18:\*  
118: gb-gss19:\*  
119: em-gss13:\*  
120: gb-gss20:\*  
121: gb-gss21:\*  
122: gb-gss22:\*  
123: gb-gss23:\*  
124: gb-gss24:\*  
125: em-gss14:\*  
126: em-gss15:\*  
127: em-gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	52.6	4.4	638	22	AM398687	AM398687 EST309187
2	47.6	4.0	351	23	AM596452	AM596452 s112e12.y
3	45.2	3.8	411	39	T42189	T42189 5452 Lambda
4	40.4	3.4	457	36	C71727	C71727 C71727 Rice
5	39.6	3.3	530	14	A1993966	A1993966 701494994
6	38.2	3.2	536	14	A1997116	A1997116 701552128
7	37.4	3.1	361	19	AV549266	AV549266 AV549266
8	37.4	3.1	558	19	AV560872	AV560872 AV560872
9	37	3.1	439	14	AL370502	AL370502 MCBAS8B12
10	36	3.0	280	30	BB345415	BB345415 BB345415
11	36	3.0	382	114	A2068859	A2068859 RPCI-23-4
12	35.8	3.0	324	22	BB535242	BB535242 BB535242
13	35.6	3.0	525	111	A0824936	A0824936 HS_5283_A
14	35.4	3.0	266	32	BB552509	BB552509 BB552509
15	35.4	3.0	286	30	BB373760	BB373760 BB373760
16	35.4	3.0	318	28	BB203673	BB203673 BB203673
17	35.4	3.0	445	40	W90867	W90867 m179e05.r1
18	35	2.9	925	121	CNS0091P	AL053013 Drosophila
19	33.8	2.8	442	2	AA272333	AA272333 vB62d02.r
20	33.8	2.8	712	88	AQ047433	AQ047433 CLM-1a3-u
21	33.4	2.8	433	23	AM582990	AM582990 1a08g11.y
22	33.4	2.8	463	36	D48678	D48678 R1CIS15048A
23	33.4	2.8	1101	121	CNS00EHN	AL069204 Drosophila
24	33.2	2.8	309	29	BB325541	BB325541 BB325541
25	33.2	2.8	360	12	A1648665	A1648665 tx63g12.x
26	33.2	2.8	509	112	AQ916714	AQ916714 nbeu067E
27	33.2	2.8	1101	121	CNS016MN	AL106573 Drosophila
28	33	2.8	450	7	A1008381	A1008381 EST202832
29	33	2.8	585	97	AQ418315	AQ418315 RPCI-11-1
30	33	2.8	916	123	CNS03YBE	AL266099 Tetradodon
31	32.8	2.7	300	16	C08171	C08171 C08171 yu1i
32	32.8	2.7	360	16	AV189838	AV189838 AV189838
33	32.8	2.7	428	21	AW351529	AW351529 IL2-CT003
34	32.8	2.7	558	20	AW216289	AW216289 687047E01
35	32.8	2.7	668	25	AM850622	AM850622 IL3-CT021
36	32.6	2.7	266	32	BB540703	BB540703 BB540703
37	32.6	2.7	300	36	C36520	C36520 C36520 yu1i
38	32.6	2.7	312	28	BB378305	BB378305 BB378305
39	32.6	2.7	322	28	BB233297	BB233297 BB233297
40	32.6	2.7	324	36	D71830	D71830 CELK082C8R
41	32.6	2.7	341	28	BB210509	BB210509 BB210509
42	32.6	2.7	344	28	BB210443	BB210443 BB210443
43	32.6	2.7	411	24	AM829046	AM829046 ra33d03.y
44	32.6	2.7	521	91	AQ305162	AQ305162 HS_2019-B
45	32.6	2.7	719	33	BE038450	BE038450 AA17F09.A

## ALIGNMENTS

LOCUS	AM398687	638 bp	mRNA	EST	07-FEB-2000
DEFINITION	EST309187 L. pennellii trichome, Cornell University Lycopersicon pennellii cDNA clone cLPT4016 5', mRNA sequence.				
ACCESSION	AM398687				
VERSION	AM398687.1 GI:6917157				
KEYWORDS	EST.				
SOURCE	Lycopersicon pennellii.				
ORGANISM	Lycopersicon pennellii.				
REFERENCE	Alcala,J., Vredalov,J., White,R., Matern,A.L., Lakey,J., Holt,I.E., Liang,F., Hansen,T.S., Upton,J., Rinning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Mierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J. Generation of ESTs from wild tomato (Lycopersicon pennellii) trichomes Unpublished (1999) Contact: David Fritsch Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfritsch@CLEMSON.EDU 5 prime sequence.				
AUTHORS	1. (bases 1 to 638)				
TITLE	Alcala,J., Vredalov,J., White,R., Matern,A.L., Lakey,J., Holt,I.E., Liang,F., Hansen,T.S., Upton,J., Rinning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Mierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J. Generation of ESTs from wild tomato (Lycopersicon pennellii) trichomes Unpublished (1999) Contact: David Fritsch Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfritsch@CLEMSON.EDU 5 prime sequence.				
JOURNAL	Unpublished (1999) Contact: David Fritsch Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfritsch@CLEMSON.EDU 5 prime sequence.				
COMMENT	Unpublished (1999) Contact: David Fritsch Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfritsch@CLEMSON.EDU 5 prime sequence.				
FEATURES	Location/Qualifiers 1..638 /organism="Lycopersicon pennellii" /db_xref="taxon:28526" /clone="cLPT4016" /clone_1b="L. pennellii trichome, Cornell University" /issue_1b="trichome" /issue_type="trichome" /dev_stage="mixed stages" /lab_host="SOLR" /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Leaves of various stages were shaken in liquid nitrogen, shearing off trichomes. This procedure yielded a mixture of cells highly enriched for trichomes, with minor contamination by other types of leaf cells." 126 c 164 g 169 t				
BASE COUNT	179 a 126 c 164 g 169 t				
ORIGIN	179 a 126 c 164 g 169 t				
Query Match	4.4% Score 52.6; DB 22; Length 638;				
Best Local Similarity	45.5% Pred. No. 1.7e-05;				
Matches 187; Conservative	0; Mismatches 224; Indels 0; Gaps 0;				
OY	582	AGCAACACGAGCGTATGCTATGCTAGCGTCAAGTGCATCGCGGCGGCGTCCG	641		
DB	228	AACACCCACGAGCAGTATTTTCATCGGACCGCTTATGTTAGTGGCTTTAGC	287		
OY	642	TGGGACAGCGCGGCGGACGATGCGTGCATATCGCGTTCAGCGAAGCGTTGT	701		
DB	288	TGGAACAGCGGCTAGGCGTGCATGAGCTTCGATCTTAATATACATGATTTACT	347		
OY	702	AGCCAGTCACAAAACCGCATTTGACATCATATATGCTGGCAGTTTCAGCACGCTT	761		
DB	348	ATCCAGGCTTAAGGACCATATGATTTGCTATAGTACGGAGAGTCATTAAGAAATTT	407		
OY	762	GCAAGACGACGACGCTGCTAAAGTGCCGCGCATGCCAAGTTTACTCAAAATAAGCA	821		
DB	408	GGAGGCGTGTGTTCAAGCTTTATTTGAACCAAGAAAGCAATTAAGAAATTTACAG	467		
OY	822	AGTTACAGCTTTTACACACCAATTACAGGGGACATTCGCGACATTTAAGTGTACCGC	881		
DB	468	AGTTCAACATCTTACGCTGATTTAGAGGGAGACCTCAGGCTGAAGATGATGAGTTTA	527		
OY	882	GATTGTTGACCGCTTGATCCACACGACGATGGGCGTCCACGATGACGCGCTT	941		
DB	528	GATTATGTCAATCGATGACACCTACTCCAGCAGTTTGTGGTATCTTACAGAAATGACG	587		



Oy	942	GATATACATTGGAGCACCATGAGAACACACCGTCGTGGTTGCAGCTCC	992
Db	588	GCCITTTATTTCAGAACTGAATCTTTTGACCGAAGAATGATGCTGCTCC	638
RESULT	2		
AM596452			
LOCUS			
DEFINITION	AM596452	351 bp mRNA EST	22-MAR-2000
ACCESSION	sJ12e12.v1	Gm-cl032 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:	
KEYWORD	Gm-cl032-1703	5' similar to TR:Q9ZPC0 Q9ZPC0 ISOCHORISMATE SYNTHASE	
VERSION		, mRNA sequence.	
REVISION	Am596452		
SOURCE	AM596452.1	GI:7283850	
ORGANISM	EST.		
	soybean.		
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
AUTHORS	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;		
	Fabales; Fabaceae; Papilionoideae; glycine.		
TITLE	1 (bases 1 to 351)		
JOURNAL	Shoemaker,R., Keim,P., Vodkin,L., Erpellding,J., Coryell,V., Khanna		
COMMENT	A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,		
	Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers		
	Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk		
	R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,H., McCann		
	R., Waterston,R. and Wilson,R.		
	Public Soybean EST Project		
	Unpublished (1999)		
	Contact: Shoemaker R/Public Soybean EST Project		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	This clone is available through: Genome Systems, Inc. 4633 World		
	Parkway Circle St. Louis, Missouri 63134 For further information		
	call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)		
	427-3324 or contact: clones@genomesystems.com or		
	info@genomesystems.com web site: www.genomesystems.com		
	High quality sequence stop: 195.		
FEATURES	location/qualifiers		
SOURCE	1..351		
	/organism="Glycine max"		
	/db_xref="taxon:3847"		
	/clone="GENOME SYSTEMS CLONE ID: Gm-cl032-1703"		
	/clone_lib="Gm-cl032"		
	/tissue_type="Cotyledons of 8-day-old 'Williams'		
	'seedlings"		
	/lab_host="DH10B"		
	/note="Vector: pluescript II XR; Site.1: EcoRI; Site.2:		
	XhoI; This cDNA library was constructed from mRNA isolated		
	from cotyledons of 8-day-old 'Williams' seedlings which		
	were propagated on paper towels with distilled water for		
	3 days (etiolated), then greenhouse grown for 5 days in		
	popping soil. The cotyledons were flash-frozen in liquid		
	nitrogen. Stratagene's cDNA Synthesis kit (catalog number		
	200401) was used to synthesize the cDNA. First-strand		
	synthesis was performed with 5-methyl dCTP, hence the		
	ligated cDNA was hemimethylated. A modification of		
	Stratagene's first-strand synthesis primer was used. An		
	'anchor' nucleotide (V-A, C, or G) was added to the 3' end		
	of the primer [GAGACGAGACGAGACGACTACTGCTGCA(7T)18V] to		
	anchor the primer at the 5' end of the poly(A) tract.		
	After second strand synthesis, the cDNA ends were filled		
	in with cloned Pfu DNA, ligated to EcoRI adapters and		
	subsequently phosphorylated. The cDNA was then		
	precipitated and redissolved in sterile, RNase-		
	DNase-free water. The XhoI site within the first-strand		
	synthesis primer was then restricted by digestion with		
	XhoI from Promega (40U/ul); all XhoI sites in the cDNA		
	would be protected by their hemimethylated status. The		
	cDNA constructs were size-fractionated with a 500 bp		

Query Match	Best Local Similarity	Score	DB	Length
Matches 160; Conservative	51.4%;	4.0%;	Pred. No. 0.00053;	351;
	0;	Mismatches 145;	Indels 6;	Gaps 2;
OY 826	CAGCACTTTGTACACACCAATTACAGGGGACATTCGGGCACATTTTAAGTGTACCCGCATT	885		
Db 1	CAACATTTATATTCATCATTTATCTGGGAGGTTAAGACGTGAAGAAGATCTTGAATTT	60		
OY 886	GTTGACCCGCTTGACATCCACACACGACACTGGGTGGCGTCGCCACGTAAGCGCCCTGTAT	945		
Db 61	TTTGATCATCTCTTCCACCCGAGTCACACATTTTGTGGGCTTCCACAGAGAGGACCAACTT	120		
OY 946	TACATTTGGACCCATGAGAGACACACCTGCTGGCTGTTTGGCAGGTCTATTTGGCTATTTT	1005		
Db 121	TTAATTTGACAGATGACGAAGATATTGTGATCGAGGAGATGTATGTCTGGACCTGTGGTTG	180		
OY 1006	ACCCGACATTAATAGTGGGAAATTTGTGTGGCATCCCTTCCATGATATGTAATCAACG	1065		
Db 181	GGAGGAGGAGAGAGAGCA---TTTCCGCTTGGCATCAGCATGAGCATTTGGTGG---AAAG	234		
OY 1066	CAGCAGCAGACAACTTTATTTGCTGCTGCTCCGGGATTTGGCTGACCTCCGATGCCACAA	1125		
Db 235	GATCTTGCTGCTTATATATATGCTGGAGAGGATGTGGAAGACCAATCTTACTTG	294		
OY 1126	GAATATGAGA 1136			
Db 295	GAGTGGATGA 305			
RESULT 3				
LOCUS T42189	411 bp	mRNA	EST	07-JAN-1998
DEFINITION 5452 lambda-PR2 Arabidopsis thaliana cDNA clone 111A1217, mRNA sequence.				
ACCESSION T42189				
VERSION T42189.1	GI:2757580			
KEYWORDS EST.				
SOURCE Arabidopsis thaliana				
ORGANISM Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE 1 (bases 1 to 411)				
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,E., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M., Retzel				
TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones				
JOURNAL MEDLINE 95148729				
COMMENT On Jan 7, 1998 this sequence version replaced gi:933036.				
CONTACT: Thomas Newman				
MSU-DOE Plant Research Laboratory				
Michigan State University				
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.				
Lansing, MI				
Tel: 517-353-0854				
Fax: 517-353-9168				
Email: 22313c@msu.edu				

FEATURES	Seq primer: 17.	Location/Qualifiers
SOURCE	1. 411	/organism="Arabidopsis thaliana"
		/strain="var columbiana"
		/db_xref="taxon:3702"
		/clone="11A12P7"
		/clone_lib="lambda-PRL2"
		/note="Vector: lambda zip-lox; Site_1: Sal; Site_2: Not; Lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRU's lambda zip-lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dt primed cDNA."
BASE COUNT	124 a 61 c 107 g 103 t 16 others	
ORIGIN		
Query Match	3.8%; Score 45.2; DB 39; Length 411;	
Best Local Similarity	50.8%; Pred. No. 0.0032;	
Matches 101; Conservative	0; Mismatches 98; Indels 0; Gaps 0;	
OY	822 ACTTCAGCATTTGTATACACACCAATTACAGGGAGACATTGGCGGACATTTAAAGTGTGACCGC 881	
DB	86 ATGTCACACNCTATATNTCTCAATTGGCAGGAGACTTACGAAGAAGATGATGAGTATA 145	
OY	882 GATTCTTGACCCCTTGCAATCCAAACACAGCATGGGTGGCGTCCACGATGAAGCGGCTT 941	
DB	146 AATATTTGGCTGCTGCAATCCAAATCCACCTTTTGTGGCTTCCAGCAGGAAGACAAAG 205	
OY	942 GTATTACATTGGGACCATCATGAGAACAACCTCGTGGCTGTTTGGAGGTCATTGGGCTA 1001	
DB	206 GCTTTTAAATTAAGAGATGATGATCATGTCATGAGAGGAAATNTATGCGGGACCTATTGGATT 265	
OY	1002 TTTTACCCGAGATAATAGT 1020	
DB	266 TTTTGGTGGCGAGAGAGAT 284	
RESULT 4		
C71727	457 bp mRNA	EST 22-SEP-1997
LOCUS	C71727 Rice panicle at flowering stage Oryza sativa cDNA clone	
DEFINITION	E0143.1A, mRNA sequence.	
ACCESSION	C71727	
VERSION	C71727.1	GI:2427264
KEYWORDS	EST.	
SOURCE	Oryza sativa.	
ORGANISM	Oryza sativa	
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.	
AUTHORS	1 (bases 1 to 457)	
TITLE	Sasaki,T. and Yamamoto,K.	
JOURNAL	Rice cDNA from panicle at flowering stage	
COMMENT	Unpublished (1996)	
	Contact: Takuji Sasaki	
	National Institute of Agrobiological Resources	
	Rice Genome Research Program	
	2-1-2 Kannondai, Tsukuba	
	Ibaraki,	
	Japan 305	
	Tel: 0298-38-7441	
	Fax: 0298-38-7468	
	Email: tsesaki@abr.affrc.go.jp.	
FEATURES	Location/Qualifiers	
SOURCE	1. 457	/organism="Oryza sativa"
		/strain="Nipponbare"
		/db_xref="taxon:4530"

```

/cclone="E0143_1A"  
/cclone_lib="Rice panicle at flowering stage"  
/dev_stage="flowering stage"  
/note="Organ: panicle; Rice cDNA from panicle at flowering  
stage"  
  
BASE COUNT      130 a     83 c    113 g    130 t     1 others  
ORIGIN  
  
Query Match          3.4%; Score 40.4; DB 36; Length 457;  
Best Local Similarity 55.8%; Pred. No. 0.11;  
Matches   77; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
  
QY       896 TCATCCCAACACCAGCAGCTGGGTGCGTCGCCACGCATGAAGCGCCCTGATTACTTACCGCA 955  
            | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||  
DB        186 TTCATCCGCGCCAGCCAGCCGTTTGTCGTCCACCAAGAAGGACAGCCCATTCATCAACAAG 245  
            | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||  
QY       956 CCCATGAGAAAGAACACCTCGTGGCTGTTTGACAGTCCTATTGGCTATTATTACCGAGATA 1015  
            ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||  
DB        246 ATTATGAAATTTTTTGGACCTGCGTAATGATCTGGACCTGTGTTGGAGAGACGTCG 305  
            | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||  
QY       1016 ATATGTGGCAATTGTGNG 1033  
            | ||||| | |||||  
DB        306 AATGTGAGTTCCTGTG 323  
            | ||||| | |||||  
  
RESULT         5  
AI993966      530 bp      mRNA      EST      08-SEP-1999  
LOCUS         AI993966/C  
DEFINITION    701494994 A. thaliana, Ohio State clone set Arabidopsis thaliana  
                CDNA clone 701494994, mRNA sequence.  
ACCESSION     AI993966  
VERSION       AI993966  
KEYWORDS      AI993966..1 GI:5840871  
SOURCE        EST.  
ORGANISM      thale cress.  
              Arabidopsis thaliana  
              Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta:  
              Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eurosids II;  
              Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE     1 (bases 1 to 530)  
AUTHORS       Chen,Y., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D.,  
Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brzoska,P.,  
Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R.,  
Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T.,  
Policky,J., Suzuki,I.G., Argentine,C., Shah,S., Nobrigo,A., Murry,L.,  
Turner,C., Krikorian,S., Elder,L., and Hanson,D.  
Arabisdisps thaliana Gene Expression Microarray  
Unpublished (1999)  
COMMENT       Contact: David Smoller, Ph.D.  
Genome Systems, Inc., a wholly owned subsidiary of Incyte  
Pharmaceuticals, Inc.  
4633 World Parkway Circle, St. Louis, MO 63134, USA  
Tel.: 877-577-2733  
Fax: 314-627-3324  
Email: service@genomesystems.com.  
  
TITLE         JOURNAL  
JOURNALNL  
COMMENT
```

OY	497	TGCATCCAACTGCCACTGTTTGTGGCGCTTCACGACAGAGAAGCAAGCGCTTTGATTGAAG	438
OY	936	CCCATGAGAAGACACCCTCGTGCTGTTTGCAGGTCTATTGGCTATTTTACCAGAGATA	1015
DB	437	AGATGATCATTCATCATGAGGAATGTAMCGGACCATATTGATTTTTTGGTGGCAGG	378
OY	1016	ATAGTG 1021	
DB	377	AGAGTG 372	
RESULT#	6		
LOCUS	A1997116/c		
DEFINITION	701552128 A. thaliana, Columbia Col-0, root-2 Arabidopsis thaliana		
ACCESSION	CDNA clone 701552128, mRNA sequence.		
VERSION	A1997116		
KEYWORDS	A1997116.1 GI:5844021		
SOURCE	EST.		
ORGANISM	thale cress. Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.		
AUTHORS	Chen,J., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D., Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brzoska,P., Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R., Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T., Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobriva,A., Murry,L., Turner,C., Krikorian,S., Elder,L. and Hanson,D. Arabidopsis thaliana Gene Expression Microarray Unpublished (1999) Contact: David Smoller, Ph.D. Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc. 4633 World Parkway Circle, St. Louis, MO 63134, USA Tel.: 877-577-2733 Fax: 314-427-3324 Email: service@genomesystems.com.		
TITLE	JOURNAL		
COMMENT			
FEATURES			
source	Location/Qualifiers		
	1..536		
	/organism="Arabidopsis thaliana"		
	/cultivar="Columbia Col-0"		
	/db_xref="taxon:3702"		
	/clone="701552128"		
	/clone_1lb="A. thaliana, Columbia Col-0, root-2"		
	/tissue_type="root"		
	/dev_stage="4 - 7 weeks"		
	/note="Vector: pSPORT; Site_1: NotI; Site_2: SalI; CDNA library was derived from untreated root tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was bluntended, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."		
BASE COUNT	159 a 119 c 91 g 145 t 22 others		
ORIGIN			
Query Match	3.2%; Score 38.2; DB 14; Length 536;		
Best Local Similarity	45.4%; Pred. NO. 0.57; Indels 0; Gaps 0;		
Matches	94; Conservative 0; Mismatches 113;		
OY	62	CACCTGAAACAACACTGAATTTGGCGCTTACTTGGCACACCCCGTAGATGATTACTTACGTT	121
DB	393	CATCTCANNNNNNNNNNNNNNNNNNNNMTGTCCTGCACTAGTATGATGATGCTTTGACGC	334
OY	122	TTCGCAATGGCCCATCGCTACGCGAAAAAGCGCTCAGGCAATTAAGGTGCGTTGTTT	181

Db	333	GGTGGCCATGTCACCTTCTCTCCAAAACTGCCAATTAAGACATGACTGTGTGCA	274
Qy	182	TTTGTCGCGAGTCATTGTGATGAACAAGATACCCGACGTCTGTAATTGATGGCGGTTT	241
Db	273	TTTGTGTGGATGCACCATCTGAAGCAGTGTGTGCGGTGTGACATGTGCGCGATGCA	214
Qy	242	GGTTGTGTCGCCAAGTGAATGATGACCA	268
Db	213	TTTCTTGTGTTGAAGAAGATTGAACA	187
RESULT	7		
LOCUS	AV549266	361 bp	EST
DEFINITION	AV549266 Arabidopsis thaliana roots Columbia Arabidopsis thaliana		23-JUL-2000
ACCESSION	CDNA clone R205e08 5', mRNA sequence.		
VERSION	AV549266		
KEYWORDS	AV549266.1 GI:8720679		
SOURCE	EST.		
ORGANISM	thale cress.		
	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 1 to 361)		
AUTHORS	Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.		
TITLE	A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries		
JOURNAL	DNA Res. 7, 175-180 (2000)		
COMMENT	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.		
FEATURES			
source	Location/Qualifiers		
	1..361		
	/organism="Arabidopsis thaliana"		
	/strain="Columbia"		
	/db_xref="taxon:3702"		
	/clone="R205e08"		
	/clone_lib="Arabidopsis thaliana roots Columbia"		
	/tissue_type="roots"		
	/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	103 a 72 c 75 g 111 t		
ORIGIN			
Query Match	3.1%; Score 37.4; DB 19; Length 361;		
Best Local Similarity	48.8%; Pred. No. 0.89;		
Matches 101; conservative	0; Mismatches 106; Indels 0; Gaps 0;		
Qy	62 CACTTGAACAACGTAATTTGGCGCTTACTTTGGACACCCGCTGATGATTAACGTT	121	
Db	22 CACTCAGCAACAACAACAACAACCAATGTTCTGCACTAAGTATGATGCTTGAATC	81	
Qy	122 TTGGCATTTGGCCCATTCGTCAGCGAATAACGGCTCAGCATTTACAAGTGGCGTTT	181	
Db	82 GGTGTGCGCATGTGACTCTCTCTCCAAAACGCGAATAAAGAAGATGAGCTGTGTGA	141	
Qy	182 TTGGTGGCAGTCATTTGATGAACAAGATTAACCCGACGTGTAATTGATGGCGGTTT	241	
Db	142 TTTGTGTGATGCACCATCTGAAGCAGTGTGTGTGCGGTGTGACATGTGCGCGATGCA	201	
Qy	242 GGTTTGTCCCGAAGTATGTGACCA	268	
Db	202 TTTCTTGTGTTGAAGAAGATTGAACA	228	
RESULT	8		
LOCUS	AV560872	558 bp	mRNA
			EST
			23-JUL-2000

[illegible]

BASE COUNT	FEATURES	TITLE
ORIGIN	SOURCE	JOURNAL
		COMMENT
137 a	1. 439	Medicago truncatula ESTs from nitrogen-starved roots
74 C	/organism="Medicago truncatula"	Unpublished (2000)
130 G	/cultivar="Jemalong"	Contact: Genoscope
98 T	/db_xref="taxon:3680"	Genoscope - Centre National de Sequencage
	/clone="MtBA38B12"	Bp 191 91006 EVRY cedex - France
	/clone.lib="MtBA"	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
	/tissue.type="root tips"	Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
	/dev_stage="harvested after 3 days of N-starvation"	Biologie Moleculaire des Relations Plantes-Microorganismes,
	/note="Vector: pBluescript PSK; Site_1: EcoRI; Site_2: XhoI; Plants were grown in an aeroponic chamber for 14 days on nitrogen-rich medium followed by 3 days on N-free medium. RNA was extracted from root tips (1-3 cm). cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zapR vector from StrataGene and packaged using GigaPack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExsacII helper phage and propagated in SGR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."	CNRS-INRA, Bp 27 3132 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).
	Location/Qualifiers	

	Query Match	3.1%	Score 37	DB 14	Length 439	
	Best Local Similarity	53.9%	Pred. No. 1.3			
	Matches 76	Conservative	0	Mismatches 65	Indels 0	Gaps 0
OY	54 CTATCGGCACACTGTAAGAACACATGAATTGGCCCTTACTTTGGCAGACC GCCGTGATCATAC 113   					
Dd	225 CGATAGGAGAAGTGTGAAGGTGTGAAGATGATCATCCCTTAGTCATAAACA CATTTGGTCAGAG 284   					
OY	114 TTTAGCTTTGGCATGTGGCGCAATGCTACGCCAAAAAAGGCGTCAGGCAATTC AAGTGAC 173   					
Dd	285 TTTCGGGTTTTTCCGAGTGTGAGTCCA AAACGGAGATCAATTC AAGCTGTATGATGA 344   					
OY	174 GATTGTTTTTGGTGGCGCAGTC 194   					
Dd	345 AGCTATGTTTGGTGTCTCCGCGC 365   					
RESULT 10						
BB345415						
LOCUS	BB345415	280 bp	mRNA	EST	12-JUL-2000	
DEFINITION	BB345415 RIKEN full-length enriched, 10 days neonate cerebellum Musculus CDNA clone B93J0056M08 3', mRNA sequence.					
ACCESSION	BB345415					
VERSION	BB345415.1	GI:9057198				
KEYWORDS	EST.					
SOURCE	house mouse.					
ORGANISM	Mus musculus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
AUTHORS	1 (bases 1 to 280) Konno,H., Altawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Horii,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sekai,C., Sato,K., Shibata,K., Shibata,					



RESULT 12				
LOCUS	BB535242/c			
DEFINITION	BB535242 261 bp mRNA EST 31-JUL-2000			
ACCESSION	BB535242 RIKEN full-length enriched, 0 day neonate lung Mus musculus cDNA clone E030042118 3' similar to U06421 Murine trombospondin 2 mRNA, exon 22 (3'UTR), mRNA sequence.			
VERSION	BB535242			
KEYWORDS	BB535242.1 GI:9590741			
SOURCE	EST.			
ORGANISM	house mouse.			
REFERENCE	Mus musculus			
AUTHORS	Eukaryote: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 261)			
	Kono, H., Atawak, K., Akahira, S., Akiyama, J., Arakawa, T., Carrinci, P., Endo, T., Fukuda, S., Fukunishi, T., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Iizawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, C., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogaue, T., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.			
	RIKEN Mouse ESTs (Kono, H., et al.)			
	Unpublished (2000)			
	Contact: Yoshihide Hayashizaki			

```

FEATURES
SOURCE
    Location/Qualifiers
    1. .261
       /organism="Mus musculus"
       /db_xref="taxon:10090"
       /clone="E030042118"
       /clone_id="RIKEN full-length enriched, 0 day neonate
       Lung"
       /tissue_type="lung"
       /dev_stage="0 day neonate"
       /lab_host="DH10B"
       /note="Site_1: Salt; Site_2: BamHI; cDNA library was
       prepared and sequenced in Mouse Genome Encyclopedia
       Project of Genome Exploration Research Group in Riken
       Genomic Sciences Center and Genome Science Laboratory in
       RIKEN. Division of Experimental Animal Research in Riken
       contributed to prepare mouse tissues. 1st strand cDNA was
       plimed with a primer [5',
       GAGGAGACAGACGGCCGCCCACTGTCAGCTTTTTCCTTTTTCCTTTCN 3'], cDNA was
       prepared by using trehalose thermo-activated reverse

```

[illegible]



TITLE  
JOURNAL  
COMMENT

Search completed: November 25, 2000, 01:23:12  
Job time: 24306 sec

RIKEN Mouse ESTs (Konno, H., et al.)  
Unpublished (2000)  
Contact: Yoshitake Hayashizaki  
Genome Exploration Research Group, Life Science Tsukuba Center,  
Genome Science Laboratory  
The Institute of Physical and Chemical Research (RIKEN), Genomic  
Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: +81-298-36-9013  
Fax: +81-298-36-9098  
Email: genome-res@rct.riken.go.jp  
URL: http://genome.rct.riken.go.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,  
N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Thermostabilization and thermocatalysis of thermostable enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,  
Y., and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (http://genome.rct.riken.go.jp) for  
further details.

FEATURES  
Source

Location/Qualifiers  
1..286  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="C130071N21"  
/clone\_1lb="RIKEN full-length enriched, 16 days embryo  
head"  
/sex="mixed"  
/tissue\_type="head"  
/dev\_stage="16 days embryo"  
/lab\_host="DH10B"  
/note="Site 1: SalI; Site 2: BamHI: cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGAGATCCCAAGACTCTTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGAGATCCGAGTATTAATTAATCCGCCGCCGCC 3']. cDNA  
was cloned into the XhoI and BamHI sites. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda  
FLC I"

BASE COUNT 90 a 40 c 39 g 117 t  
ORIGIN

Query Match 3.0%; Score 35.4; DB 30; Length 286;  
Best Local Similarity 56.4%; Pred. No. 3.5;  
Matches 66; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 1078 ACTTTATTTGCGTGGCGGANTGCTGACTCCGATGCCACAGAAATATGAAGAA 1137  
DB 273 AATTATTTTGTCTGTATATCTTTGTATATAAATGAAAGCTTACTGATTAAGGGTT 214  
QY 1138 ACTGGGTGAATTTGAACCATGCGCAATGTTTAAGACTACATCATGTGAA 1194  
DB 213 TTAGTGTGAATTTTACCCATGACCAATATAATATTTTACCAATCTTGGAA 157





**THIS PAGE BLANK (user)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2000, 01:23:18 ; Search time 55.47 Seconds  
(without alignments)  
244.726 Million cell updates/sec

Title: US-08-774-104a-2  
Perfect score: 2009  
Sequence: 1 MMTYHETRALAQSLLQQLYA.....ETGLKFERMRLDYNHIVE 397

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq.36:\*

- 1: /SID6/gcgdata/geneseq/AA1980.DAT:\*
- 2: /SID6/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SID6/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SID6/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SID6/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SID6/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SID6/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SID6/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SID6/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SID6/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SID6/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SID6/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SID6/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SID6/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SID6/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SID6/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SID6/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SID6/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SID6/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SID6/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SID6/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2009	100.0	397	19	W69162 Human neutral sph
2	360	17.9	580	20	Y50340 C. roseus ICS prot
3	344	17.1	591	20	Y50335 E. coli entC prote
4	329.5	16.4	291	20	Y33698 E. coli entC prote
5	302.5	15.1	391	20	Y50336 P. fluorescens ICS
6	297	14.8	331	20	Y50338 P. fluorescens ICS
7	209	10.4	491	20	Y41302 Tobacco anthranila
8	195.5	9.7	603	20	Y42112 Corn anthranilate
9	195	9.7	616	20	Y41301 Tobacco anthranila
10	189.5	9.4	606	20	W93814 Rice anthranilate
11	186.5	9.3	446	20	Y42113 Corn anthranilate
12	184.5	9.2	615	18	W26685 Maize C28 allele a

13	183.5	9.1	615	18	W26686
14	181.5	9.0	577	20	W93815
15	181.5	9.0	577	20	W93810
16	170	8.5	518	8	P70506
17	153	7.6	401	18	W55536
18	153	7.6	567	18	W55703
19	152	7.6	430	18	W55418
20	150	7.5	193	9	P80877
21	150	7.5	262	5	P40023
22	150	7.5	405	4	P30202
23	149	7.4	559	19	W98537
24	145	7.2	225	8	P70359
25	145	7.2	225	8	P70061
26	142	7.1	406	4	P30203
27	140	7.0	609	4	P30206
28	137	6.8	529	20	Y29293
29	107.5	5.4	1224	18	W26735
30	97	4.8	2532	20	Y33729
31	96.5	4.8	1346	21	Y77195
32	96.5	4.8	1346	21	Y67204
33	96.5	4.8	1219	21	Y77180
34	96	4.8	1224	20	W36851
35	95.5	4.8	600	20	W82660
36	94.5	4.7	1374	19	W72225
37	94.5	4.7	1384	19	W72224
38	94.5	4.7	1386	19	W72117
39	94.5	4.7	1396	19	W72039
40	94	4.7	1092	19	W41602
41	94	4.7	3739	21	Y77193
42	94	4.7	3739	21	Y77201
43	94	4.7	3739	21	Y67202
44	93.5	4.7	991	21	Y83171
45	93.5	4.7	991	21	Y70120

## ALIGNMENTS

Maize anthranilate  
Rice ASA first iso  
Rice anthranilate  
Type enzyme. P70  
H. pylori ORF 07ap  
H. pylori ORF 07ee  
H. pylori ORF hp2e  
Hook region #8 con  
Fusion protein con  
Sequence encoded b  
H. pylori GPO 148  
Sequence encoded b  
Sequence of human  
Sequence encoded b  
Sequence encoded b  
Deduced amino acid  
Staphylococcus car  
Photobacterium lum  
S. venezuelae macr  
Nardoniidae syntha  
S. venezuelae pik  
Protein involved i  
Cauliflower L-gala  
HSV-2 strain SB5 C  
HSV-2 strain SB5 C  
HSV-2 strain SB5 C  
Staphylococcus epl  
S. venezuelae macr  
S. venezuelae pik  
Nardoniidae syntha  
Cell wall protein  
Staph. epidermidis

RESULT 1	
W69162	W69162 standard; Protein; 397 AA.
ID	
AC	W69162;
DT	20-OCT-1998 (first entry)
XX	
DE	Human neutral sphingomyelinase.
XX	
KW	Human: neutral sphingomyelinase; N-Smase; diagnosis; Crohn's disease;
KW	obesity; diabetes; Alzheimer's disease.
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 59 /note="encoded by GTTCTT"
XX	
PN	W09828445-A1.
PD	02-JUL-1998.
XX	
PF	23-DEC-1997; 97MO-US24051.
XX	
PR	24-DEC-1996; 96US-0774104.
XX	
PA	(UYJO ) UNIV JOHNS HOPKINS.
XX	
PI	Chatterjee S;
XX	
DR	WPI: 1998-377673/32.
XX	
XX	N-PSDB: V41196.
PT	Human neutral sphingomyelinase - used to, e.g. treat N-Smase

PT associated disorders, e.g. Crohn's disease, obesity, diabetes, and  
PT Alzheimer's disease  
PS Claim 19; Page 29-30; 47pp; English.  
XX  
CC The present sequence represents human neutral sphingomyelinase (N-SMase).  
CC A host cell containing a vector comprising a nucleotide sequence  
CC encoding N-SMase can be used to produce N-SMase. N-SMase can be used in  
CC a method for identifying a compound useful in the diagnosis or treatment  
CC of a human neutral sphingomyelinase related disorder. N-SMase, and a  
CC nucleotide sequence encoding N-SMase, can be used for modulating N-SMase  
CC activity, and for treating a disorder associated with N-SMase. The  
CC N-SMase disorders that can be treated with the nucleotide sequence or  
CC N-SMase, is an inflammatory disorder, arthritis, osteoarthritis, Crohn's  
CC disease, obesity, diabetes, cirrhosis, susceptible tumours, central  
CC nervous system disorder, vascular stenosis, arterial occlusion arising  
CC from plaque formation, cardiac disease where LV dysfunction occurs,  
CC hypercholesterolaemia, cholesterol ester storage disorder, renal failure,  
CC HIV infection, depression, schizophrenia, neurodegeneration, and  
CC Alzheimer's disease. An antibody against N-SMase can be used to reduce  
CC tumour necrosis factor alpha (TNF-alpha) induced apoptosis of mammalian  
CC cells.  
XX  
SQ Sequence 397 AA:  
  
Query Match 100.0%; Score 2009; DB 19; Length 397;  
Best Local Similarity 100.0%; Pred. No. 8.6e-181; Mismatches 0; Gaps 0;  
Matches 397; Conservative 0; Indels 0; Gaps 0;  
  
QY 1 MMTHTERRALAQSDILOQYLALETTEFGAYRTPADDTLRGIGAIATAKTAQALQGAVF 60  
DB 1 mmyhetralaqsdldqlyaaletetefgafatpaddtlrfgialataktaqalgavf 60  
QY 61 GAOSFDEQEXYEOSELMAGFWFVPEVMTIADKIRFGSDYSDFTTWIAQFYPKQPNY 120  
DB 61 gagsfdeeqysoselemagfwfvpewmtiaadkirtfgsdvdsfttcwlaqtpkqpnvt 120  
QY 121 TSHVTDEVDWIERENLIDTLAIDQTLAKVVGROOTLQLSPTLRALQIIRALAEQANTY 180  
DB 121 tshvtdevdwierenlidtlaidqlakvvgfgrqtlqsdltlrqaliralaeanqy 180  
QY 181 HVLKRDHDLTISATPERLVAMSGQIATPAVAGTSRRGTGADGADIALGEALLASOKNRI 240  
DB 181 hvltkrdhdeltsatperlvamsgqiatpaavgtsrrgtgdadialgeallaaqknri 240  
QY 241 EHOVVASITRRLODVTSKLVPAWPSLLKKNQVHLYTPITGDIATLAHLSTVATYDRLHP 300  
DB 241 ehovvasitrrlodvtsklvpawpsllkknqvhlytpitgdiaahlstvaydrlhp 300  
QY 301 TPALGVPREALYYIATHEKTPRGLFAGPIGFTADNSGEFVVGIRSMYVNOTORRATL 360  
DB 301 tpalgvpreaalyyiathektrpgrlfgapigftadnsgefvvgirsmynvnotqrratl 360  
QY 361 FAGAGIVADSDAQOEYETGLKFEPMROLKDYNHVE 397  
DB 361 fagagivadsdaqoeyetglkfepmqllkdynhve 397  
  
RESULT 2  
Y50340 Y50340 standard; Protein; 580 AA.  
XX  
AC Y50340;  
XX  
DT 14-JAN-2000 (first entry)  
XX  
DE C. roseus ICS protein.  
XX  
KM Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;  
KM pathogen inducible promoter; antipathogenic protein; toxin;  
KM antifungal protein; albumin-type protein; hypersensitive response.  
XX

OS Catharanthus roseus.  
XX  
PN WO950423-A2.  
XX  
PD 07-OCT-1999.  
XX  
PF 25-MAR-1999; 99WO-EP02176.  
XX  
PR 31-MAR-1998; 98US-0080203.  
PR 03-APR-1998; 98US-0080625.  
XX  
PA (MOGE-) MOGEN INT NV.  
PA (UYLE-) RIJCKSONIJ LEIDEN.  
PA (UYNI-) UNIV NIJMEGEN.  
PI Linhorst HJM, Verpoorte R, Verberne MC, Moreno PRH;  
PI Van Tegelen LJP, Willems GJ, Croes AF, Stulver MH, Custers J;  
PI Simons LH, Melchers LS, Bol JF;  
DR WPI; 1999-610856/52.  
XX N-PSDB; 223758.  
XX  
PT Method for inducing pathogen resistance in plants  
XX  
PS Claim 11; Page 63-64; 66pp; English.  
XX  
CC This invention describes a novel method for the induction of pathogen  
CC resistance in plants, by transformation with an expression cassette  
CC harboring a gene coding for an isochlorismate synthase (ICS). A pathogen  
CC inducible promoter can be used to drive expression of a heterologous  
CC protein. The heterologous protein used in the method of the invention is  
CC an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins,  
CC saccharide oxidase, oxalate oxidase, magainins, toxins from *Bacillus*  
CC *thuringiensis*, or antifungal proteins isolated from *Mirabilis jalapa*,  
CC *Amaranthus*, *Raphanus*, *Brassica*, *Sinapis*, *Arabidopsis*, *Dahlia*, *Citrus*,  
CC *Lathyrus*, *Cilicoria*, *Allium* seeds, *Aralia* and *Impatiens* and albumin-type  
CC proteins, such as thionine, napin, barley trypsin inhibitor, cereal  
CC gliadin and wheat-alpha-amylase, or a protein that can induce a  
CC hypersensitive response, such as Cf, Bs3 and Po proteins from tomato and  
CC N-protein from tobacco. This sequence represents the *Catharanthus roseus*  
CC isochlorismate synthase ICS protein which is described in the method of  
CC the invention.  
XX  
SQ Sequence 580 AA:  
  
Query Match 17.9%; Score 360; DB 20; Length 580;  
Best Local Similarity 29.6%; Pred. No. 1.6e-25; Mismatches 107; Conservative 65; Indels 32; Gaps 10;  
Matches 107; Conservative 65; Indels 32; Gaps 10;  
  
QY 60 FGAQSEDEQEXY--POSELMAGFWF-VPEV-----AVTIAAD-----K 93  
DB 216 ygaltrfdarphlapewkafgsfymypqvefdelhgssmlaetvwdnaalsltqgalvr 275  
QY 94 ITFGSDTVSDFTTWIAQFYPKQPNYVTSHTVDEVDWIERENLIDTLAIDQTLAKVVF 152  
DB 276 lqtlmeqysvstskrlrgdv-shstlsvskanlpdrctswdltlrvleegnkyspltkvvl 334  
QY 153 GRQ-OTLOISDTLRALQIIRALAEQANTYHVVYLRKHD-ELFTSATPERLVAMSGQIATA 210  
DB 335 arrsgvlttsdldiplawssfkadgkdayqfclqheapafignpbeqlfgdqqltvise 394  
QY 211 AVAGTSRRGTGADGADGADIALGEALLASOKNRIEHOVVASITRRLODVTSKLVPAWPSLLK 270  
DB 395 alaaatrargesdldlqnahdlfsspkdnhealvrenrtqldalcsvecepmksvyrk 454  
QY 271 NKQOVHLYTPITGDIATLAHLSTVATYDRLHPPALGVPREALYYIATHEKTPRGLFAGP 330  
DB 455 lkrighlyarfaragtrrseddefklsslhprpavegffmederkfiaeemdriglyagp 514  
QY 331 IGFPTADNSGEFVVGIRSMYVNOTORRATLRFAGAGIVADSDAQOEYETGLKFEPMROL 390  
DB 515 vgfifgags-dfsvygtirsallgk-dagaliyvglyvvegsdpalewgelelkaasqfmlm 572

```
QY 391 K 391
Db 573 K 573

RESULT 3
Y50335
ID Y50335 standard; Protein; 391 AA.
AC Y50335;
DE 14-JAN-2000 (first entry)
XX
XX E. coli entc protein.
XX
XX Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;
KM pathogen inducible promoter; antipathogenic protein; toxin; entc;
KM antifungal protein; albumin-type protein; hypersensitive response.
XX
OS Escherichia coli.
XX
XX WO9950423-A2.
XX
XX 07-OCT-1999.
XX
XX 25-MAR-1999; 99WO-EP02176.
XX
XX 31-MAR-1998; 98US-0080203.
PR 03-APR-1998; 98US-0080625.
XX
XX (MOGE-) MOGEN INT NV.
PA (UTLE-) RIJKSUNIV LEIDEN.
PA (UYNI-) UNIV NIJMEGEN.
XX
XX Linthorst HJM, Verpoorte R, Verberne MC, Moreno PRH;
PI Van Tegelen LJP, Willems GJ, Croes AF, Stuiver MH, Custers J;
PI Simons LH, Melchers LS, Bol JF;
DR WPI: 1999-610856/52.
DR N-PSDB: 223756.
XX
XX Method for inducing pathogen resistance in plants -
PT
XX
XX Example 3; Page 51-53; 66pp; English.
XX
XX This invention describes a novel method for the induction of pathogen
CC resistance in plants, by transformation with an expression cassette
CC harboring a gene coding for an isochorismate synthase (ICS). A pathogen
CC inducible promoter can be used to drive expression of a heterologous
CC protein. The heterologous protein used in the method of the invention is
CC an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins,
CC saccharide oxidase, oxalate oxidase, magainins, toxins from Bacillus
CC thuringiensis, or antifungal proteins isolated from Mirabilis jalapa,
CC Amaryanthus, Raphanus, Brassica, Sinapis, Arabidopsis, Dahlia, Cichus,
CC Lathyrus, Clitoria, Allium seeds, Avena and Impatiens and albumin-type
CC proteins, such as thionin, napin, barley trypsin inhibitor, cereal
CC gliadin and wheat-alpha-amylase, or a protein that can induce a
CC hypersensitive response, such as Cf, Bs3 and Pc proteins from tomato and
CC N-protein from tobacco. This sequence represents the Escherichia coli
CC isochorismate synthase entc gene which is described in the method of
CC the invention.
XX
XX Sequence 391 AA;
SQ
```

```
Query Match 17.1%; Score 344; DB 20; Length 391;
Best Local Similarity 28.6%; Pred. No. 2.9e-24;
Matches 113; Conservative 65; Mismatches 153; Indels 64; Gaps 12;
```

```
QY 19 YAALETTEFGAVFATPA-----DDTLRFGIGAITAATAQAQLQGAVF-GAGSFDQEQEXP 71
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 27 yrstfscgfafdeparngsdpspfqkkaalafadakaqgikpvmvgaipfprq p 85
```

```
QY 72 OSELMAGFNFVEPMWTIAADKITFGSDIVSDFTWLAQFVPKQNTVTTSHVTDEVMKI 131
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 86 ss-----Lype-----sw-gsfarqekqasarrfr----- 111

132 ERTENDLIDPLAT-DOFLAKVVFGRQQTQLDS--DTLRALQIIRALAEQANTYHYVLRK- 186
:::|::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 112 sgslnvverqalpqqtlftqmwaraaalatcpqvkvlslrildltadaidsqyllel 171

187 -----HDEL-----FISATPERLIVAMSGGQIATAAAGTSRGTGDADIALGEA 231
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 172 laqnpvsynfhvpladgavlllgaspeelllrkdgerfssiplagsarqgdevldeagur 231

QY 233 LLASQKNRIEHOYVVAATITRLQDYVTSIAKVAAMSLSLKNKOYOHLYPTITGDIAHLSV 291
| | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 232 llaekdrhehelvpcamkevlrersselhvpsspqlltclplwlatpfdgkxansqena 291

QY 292 TAIIVRLHPTPLAGVPRRAALYYIATHEKTPRGFAGPIGFTADNDSGEFVGRSMYV 351
: | | | | | | | | : | : | : | : | : | : | : | : | : | : | : | :
Db 292 ltlacllhptpalsgfpnqaatqvlaelepfarellfgylvqcdsegnegwvlltrcakl 351

QY 352 NOTORRATLFFAGAGIVADSDAQOEXEETGLKFEPM 386
: | | | | | | | | : | : | : | : | : | : | : | : | : | : | : |
Db 352 renqvr--lffagagivpassplgewrelgvlkstlm 384

RESULT 4
Y33698
ID Y33698 standard; Protein; 271 AA.
XX
XX Y33698;
AC
XX
XX 14-JAN-2000 (first entry)
DE
XX
XX E. coli entc protein fragment.
XX
XX Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;
KM pathogen inducible promoter; antipathogenic protein; toxin; entc;
KM antifungal protein; albumin-type protein; hypersensitive response,
XX
XX Escherichia coli.
XX
XX WO9950423-A2.
XX
XX 07-OCT-1999.
XX
XX 25-MAR-1999; 99WO-EP02176.
XX
XX 31-MAR-1998; 98US-0080203.
PR 03-APR-1998; 98US-0080625.
XX
XX (MOGE-) MOGEN INT NV.
PA (UTLE-) RIJKSUNIV LEIDEN.
PA (UYNI-) UNIV NIJMEGEN.
XX
XX Linthorst HJM, Verpoorte R, Verberne MC, Moreno PRH;
PI Van Tegelen LJP, Willems GJ, Croes AF, Stuiver MH, Custers J;
PI Simons LH, Melchers LS, Bol JF;
DR WPI: 1999-610856/52.
DR N-PSDB: 223756.
XX
XX Method for inducing pathogen resistance in plants -
PT
XX
XX Example 3; Page 53-54; 66pp; English.
XX
XX This invention describes a novel method for the induction of pathogen
CC resistance in plants, by transformation with an expression cassette
CC harboring a gene coding for an isochorismate synthase (ICS). A pathogen
CC inducible promoter can be used to drive expression of a heterologous
CC protein. The heterologous protein used in the method of the invention is
CC an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins,
CC saccharide oxidase, oxalate oxidase, magainins, toxins from Bacillus
```

CC thuringiensis, or antifungal proteins isolated from *Mirabilis jalapa*,  
 CC *Amaranthus*, *Raphanus*, *Brassica*, *Sinapis*, *Arabidopsis*, *Dahlia*, *Cnicus*,  
 CC *Lathyrus*, *Clitoria*, *Allium* seeds, *Aralia* and *Impatiens* and albumin-type  
 CC proteins, such as thionine, napin, barley trypsin inhibitor, cereal  
 CC gliadin and wheat-alpha-amylase, or a protein that can induce a  
 CC hypersensitive response, such as Cf, B33 and Po proteins from tomato and  
 CC N-protein from tobacco. This sequence encodes the *Escherichia coli*  
 CC isochorismate synthase *entC* gene which is described in the method of  
 CC the invention.  
 CC  
 XX Sequence 271 AA;

Query Match 16.4%; Score 329.5; DB 20; Length 271;  
 Best Local Similarity 32.5%; Pred. No. 3.9e-23;  
 Matches 91; Conservative 46; Mismatches 124; Indels 19; Gaps 5;

QY 110 QEPKQDNVTYTHSVDEVMIEENTLIDTLAIDQLAKVFGROQLDSTLRLOI 169  
 DB 1 qalpeq-----lfegmvaraaaltatpgvd---kvvlslriditdaaidsqvl 47  
 QY 170 IRLAEO---AMRYHVVLKRHDELFIATPERIVAMSGGQIATAAAGTSRGTGADDI 226  
 DB 48 Ierliagnpvsynfihvpl-adggvlllgaspeellrkqgeritssiplqgsarrqpdevldr 106  
 QY 227 ALGELLASQKNRIHQYVVASITRLODYTTSLKVPAMPSLKNKVOHLPTITGDIA 286  
 DB 107 eagnrllasekdrhehelveqamkevlrersselhyvpsqdlitlplrlwhatpfegkan 166  
 QY 287 AHSVTAIVRLHPTPLAGVPREALYIATHEKTRPGFACIGYFTADNSGEFVYGI 346  
 DB 167 sgenatlacilphtpalsgfpbqaatqvlaeplefrelfgyivgcdeegnewvlti 226  
 QY 347 RSMVNOTORATLFEAGAGIVADSDAQEEENGKKEPM 386  
 DB 227 rcaiklengvtr--lfagagivpassplgrewetgvkistm 264

RESULT 5  
 ID Y50336 standard; Protein; 391 AA.  
 AC Y50336;  
 DE 14-JAN-2000 (first entry)  
 XX P. fluorescens ICS orfa protein.  
 KW Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;  
 KW pathogen inducible promoter; antipathogenic protein; toxin;  
 KW antifungal protein; albumin-type protein; hypersensitive response.  
 OS Pseudomonas fluorescens.  
 XX  
 XX WO950423-A2.  
 XX  
 XX 07-OCT-1999.  
 XX  
 XX 25-MAR-1999; 99WO-EP02176.  
 XX  
 XX 31-MAR-1998; 98US-0080203.  
 XX 03-APR-1998; 98US-0080625.  
 XX  
 XX (MOGE-) MOGEN INT NV.  
 XX (UYLE-) RIKSUNIV LEIDEN.  
 XX (UYNT-) UNIV NIJMEGEN.  
 XX  
 XX Linthorst HJM, Verpoorte R, Verberne MC, Moreno PRH;  
 XX Van Tegelen LJP, Mullens GJ, Groes AF, Stulver MH, Custers J;  
 XX Simons LH, Melchers LS, Bol JF;  
 XX WPI; 1999-610856/52.  
 XX  
 XX N-PSDB; Z23757.

XX  
 PT Method for inducing pathogen resistance in plants  
 XX  
 PS Example 3; Page 54-56; 66pp; English.

CC This invention describes a novel method for the induction of pathogen  
 CC resistance in plants, by transformation with an expression cassette  
 CC harboring a gene coding for an isochorismate synthase (ICS). A pathogen  
 CC inducible promoter can be used to drive expression of a heterologous  
 CC protein. The heterologous protein used in the method of the invention is  
 CC an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins,  
 CC saccharide oxidase, oxalate oxidase, magainins, toxins from *Bacillus*  
 CC *thuringiensis*, or antifungal proteins isolated from *Mirabilis jalapa*,  
 CC *Amaranthus*, *Raphanus*, *Brassica*, *Sinapis*, *Arabidopsis*, *Dahlia*, *Cnicus*,  
 CC *Lathyrus*, *Clitoria*, *Allium* seeds, *Aralia* and *Impatiens* and albumin-type  
 CC proteins, such as thionine, napin, barley trypsin inhibitor, cereal  
 CC gliadin and wheat-alpha-amylase, or a protein that can induce a  
 CC hypersensitive response, such as Cf, B33 and Po proteins from tomato and  
 CC N-protein from tobacco. This sequence represents the *Pseudomonas*  
 CC fluorescens isochorismate synthase *orfa* protein which is described in  
 CC the method of the invention.  
 CC  
 XX Sequence 391 AA;

Query Match 15.1%; Score 302.5; DB 20; Length 391;  
 Best Local Similarity 28.7%; Pred. No. 2.3e-20;  
 Matches 107; Conservative 46; Mismatches 157; Indels 63; Gaps 10;

QY 68 QEPQSELMAGFWFV---EVMVTIADKINF---GSDTVSDFTTWLAQ----- 110  
 DB 12 eevqlaevqrsfstsgdrelavqgmrlretlpaigddanslfqqlagaldfrareeqg 71  
 QY 111 -----FVPRQPNVTYTHSVTDEVDWIER-----TENLIDTLAID 144  
 DB 72 snplivgaipdipaepcl---yipenaqwrtridihaktgmslpelegknidpdeqafx 128  
 QY 145 QTLAKVFG-----ROOTLQSDTLRLAQIRALAEQANT-----YHVVLKRHD-EL 190  
 DB 129 ravehavvnfnhsdvtrkavlsyqrellifandvdvsalqhnkqpnsgyhtvmpdgctt 188  
 QY 191 FISATPERLYAMSGGQIATAAAGTSRGTGADADIALGELLASQKNRIHQYVVASIT 250  
 DB 189 ligvspelivrkqeglsalsnplagsakrmadpeadrnadalltsekdyhgfvtqdliv 248  
 QY 251 TRLODYTTSLKVPAMPSPSLKNKVOHLXPTITGDIA-AHSVTAIVRLHPTPLAGVPR 309  
 DB 249 sqigklctqlnvpqrpisltspalwhstirlegtladpavsalqlacrlnptpavcgfpt 308  
 QY 310 EALYIYATHEKTRPGFACIGYFTADNSGEFVGIKSMVNOTORATLFEAGAGIVAD 369  
 DB 309 eratrrllrfvepfergltfcmvgwcdagngewvltlrcgvttrnkvr--lfagagivea 366  
 QY 370 SDAQOEEYETGLK 382  
 DB 367 sspdsawaevqtk 379

RESULT 6  
 ID Y50338 standard; Protein; 331 AA.  
 AC Y50338;  
 DE 14-JAN-2000 (first entry)  
 XX P. fluorescens ICS orfa protein fragment.  
 KW Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;  
 KW pathogen inducible promoter; antipathogenic protein; toxin;  
 KW antifungal protein; albumin-type protein; hypersensitive response.  
 OS Pseudomonas fluorescens.

```
XX PN MO950423-A2.
XX XX
XX PD 07-OCT-1999.
XX XX
XX PF 25-MAR-1999; 99MO-EP02176.
XX PR 31-MAR-1998; 98US-0080203.
XX PR 03-APR-1998; 98US-0080625.
XX XX
XX PA (MOGE-) MOGEN INT NV.
XX PA (UYLE-) RIJSUNIV LEIDEN.
XX PA (UYNI-) UNIV NIJMEGEN.
XX PI Linthorst HJM, Verpoorte R, Verberne MC, Moreno PRH;
XX PI Van Tegelen LJP, Willems GJ, Croes AF, Stuiver MH, Custers J;
XX PI Simons LH, Melchers LS, Bol JF;
XX DR WPI: 1999-610856/52.
XX DR N-PSDB: Z23757.
XX PT
XX PS Method for inducing pathogen resistance in plants
XX XX
XX PS Example 3: Page 58-59; 66pp; English.
XX CC This invention describes a novel method for the induction of pathogen
XX CC resistance in plants, by transformation with an expression cassette
XX CC harboring a gene coding for an isochlorismate synthase (ICS). A pathogen
XX CC inducible promoter can be used to drive expression of a heterologous
XX CC protein. The heterologous protein used in the method of the invention is
XX CC an anti-pathogenic protein e.g. chitinase, glucanase, osmotin, lectins,
XX CC saccharite oxidase, oxalate oxidase, magainins, coxins from Bacillus
XX CC thuringiensis, or antifungal proteins isolated from Mirabilis jalapa,
XX CC Amarantus, Raphanus, Brassica, Sinapis, Arabidopsis, Dahlia, Nicotiana,
XX CC Lathyrus, Cicer, Allium seeds, Aralia and Impatiens and albumin-type
XX CC proteins, such as chitinase, napin, barley trypsin inhibitor, cereal
XX CC gliadin and wheat-alpha-amylase, or a protein that can induce a
XX CC hypersensitive response, such as Cf. Bs3 and Po proteins from tomato and
XX CC N-protein from tobacco. This sequence represents a Pseudomonas
XX CC fluorescens isochlorismate synthase orfa protein fragment which is
XX CC described in the method of the invention.
XX SO
XX Sequence 331 AA;
XX
Query Match 14.8%; Score 297; DB 20; Length 331;
Best Local Similarity 32.2%; Pred. No. 6e-20;
Matches 87; Conservative 43; Mismatches 134; Indels 6; Gaps 5;
QY 116 PNVTTSHTVDEVMIEKTEMLDITLADITLAKVYFGRQOTLQSLRLAQTIRAL-A 174
DB 53 pellegknipdegafkravehavvfr-hsdvkravlsqvqrellifandvdsalqhnka 111
QY 175 EQAATYHVHLKRHD-ELFISATPERLVAMSGQIATAVAGSRGTCGADIDALGALL 233
DB 112 gnpseyhfrvmpdgctclivgsPELLIVKegllslnplagsakmadpeadrnadwli 171
QY 234 ASQNRRIEHOYVVASITTRLDVTTSLKVPAMPSLKKNKOYOHLYTPITGDA-AHLSTV 292
DB 172 tsekdhgsvftvdqivsglklctqlnvprpslispalwhlstrlegtladpavsaal 231
QY 293 AIVDRHPTPLGVPREALYYIATHEKTPRGAGPIGFTTADNSEEFGVSGISMYN 352
DB 232 qlactlhpptavcgfpfctarrllirfvepfertgltgmvgcdagngewvllircgvtv 291
QY 353 OTORRATLFAAGIYVSDAQOEYEGTGLX 382
DB 292 tnkvrlfagaglyveasspdsewvqtk 319
```

```
XX AC Y41302;
XX XX
XX DT 18-JAN-2000 (first entry)
XX DE Tobacco anthranilate synthase ASAI.
XX XX
XX KW Promoter; tobacco; anthranilate synthase; tolerance; analogue; Trp;
XX KW tryptophan; transgenic plant.
XX OS Nicotiana tabacum.
XX PN US5965727-A.
XX PD 12-OCT-1999.
XX PF 31-DEC-1997; 97US-0001826.
XX PR 26-JUL-1996; 96US-0025140.
XX PR 25-JUL-1997; 97US-0937739.
XX PA (UNII ) UNIV ILLINOIS FOUND.
XX PI Withholm JM, Brotherton JE, Song H;
XX DR WPI: 1999-579943/49.
XX DR N-PSDB: Z30446.
XX PT Plant tissue culture transformation using promoters and DNA constructs
XX PS -
XX XX
XX PS Example 2: Column 41-44; 49pp; English.
XX CC This sequence corresponds to the tobacco (Nicotiana tabacum) anthranilate
XX CC synthase ASAI. The corresponding gene's promoter and DNA constructs
XX CC containing it are useful for imparting, to a plant cell, tolerance to
XX CC an amino acid analogue of Trp, or altering the Trp content in a plant
XX CC by transforming the plant cells with an expression cassette containing
XX CC the ASAI structural gene. The promoter and construct are useful for the
XX CC production of transgenic plants.
XX SO
XX Sequence 491 AA;
XX
Query Match 10.4%; Score 209; DB 20; Length 491;
Best Local Similarity 26.4%; Pred. No. 2e-11;
Matches 77; Conservative 56; Mismatches 123; Indels 36; Gaps 9;
QY 110 QFVRKQPTVTTSHYTDVEMIERTEMLDITLADITLAKVYF-----GROOTLQSL 164
DB 191 qfvpshdn-----snvtece---eykeavvk--akehllagdlfqivlsqrfertrfdp 240
QY 165 RLAQIIRLAEQANFYHVHLKRHDELFSATPERLVAMSGQIATAVAGSRGTCGAD 224
DB 241 eyrtalrv---npssymylgargclivassPELLITKkqkivnrrplagctskrkne 298
QY 225 DIALGALLASQKNRIEHOYVVASITTRLDVTT--SLKVPAMPSLKKNKOYOHLYTPIT 282
DB 299 dkrl-eellenekegaehlmiveldgrndykvtkysvkeklmlieryshvmlstvt 357
QY 293 GDIAHLSTAVDLHPTPLGVPREALYYIATHEKTPRGAGPIGFTTADNSEEFG 342
DB 358 gelqglctcwdlraalpvgvtsgapkxkamellleleptlrrgpysgfgvsgftgdm 417
QY 343 VVGIRSM-----YVNOTQR---ATLFAAGIYVSDAQOEYEGTGLX 378
DB 418 alsitlivfctacqyntmsykdankrtrewaylqagagvvasdospdehce 469
```

RESULT 7  
Y41302  
ID Y41302 standard; Protein; 491 AA.

RESULT 8  
Y42112  
ID Y42112 standard; Protein; 603 AA.

AC Y42112;  
 XX 13-DEC-1999 (first entry)  
 DT  
 XX  
 DE Corn anthranilate synthase alpha subunit protein sequence.  
 XX  
 KW Anthranilate synthase alpha subunit; tryptophan biosynthetic enzyme;  
 KW chlorogenic acid; corn; rice; soybean; wheat; tryptophan synthase;  
 KW anthranilate synthase beta subunit; herbicide; fungicide; phenotype;  
 KW pathogen attack; identification; transgenic plant; ASAS; ASBS; TSAS;  
 XX plant breeding.  
 OS  
 XX Zea mays.  
 XX  
 XX Location/Qualifiers  
 FH Key  
 FT MISC-difference 15  
 FT MISC-difference 75 /note= "unspecified"  
 FT MISC-difference 115 /note= "unspecified"  
 FT MISC-difference 115 /note= "unspecified"  
 FT  
 XX W09949058-A2.  
 XX  
 XX 30-SEP-1999.  
 XX  
 XX 19-MAR-1999; 99WO-US06046.  
 XX  
 XX 26-MAR-1998; 98US-0079386.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Vollmer SJ, Falco SC, Broglie RM, Bryan GF, Cahoon RE;  
 PI Rafalski JA;  
 XX  
 XX WPI: 1999-580451/49.  
 DR N-PSDB; 225109.  
 DR  
 XX  
 PT New isolated tryptophan biosynthetic enzyme nucleic acids, used to  
 PT produce plants with altered tryptophan levels and for developing  
 PT herbicides or fungicides -  
 XX  
 PS Claim 21: Page 55-57; 83pp; English.  
 XX  
 CC The present invention describes isolated anthranilate synthase alpha-  
 CC subunit (ASAS), anthranilate synthase beta-subunit (ASBS), and tryptophan  
 CC synthase alpha-subunit (TSAS) nucleic acids, and protein encoded by them,  
 CC obtained from corn, rice, soybean and wheat cDNA libraries. The nucleic  
 CC acid fragments may be used to create transgenic plants in which the  
 CC disclosed ASAS, ASBS or TSAS are present at higher or lower levels than  
 CC normal or in cell types or developmental stages in which they are not  
 CC normally found. This would have the effect of altering the level of  
 CC tryptophan in those cells. Manipulation of the levels of some of the  
 CC ASAS will also results in changes in the response to pathogen attack.  
 CC Because this pathway is not followed for the production of tryptophan in  
 CC higher animals, these enzymes are very good candidates for the discovery  
 CC of herbicides and fungicides. The ASAS, ASBS or TSAS can be used as  
 CC targets to facilitate design and/or identification of inhibitors of those  
 CC enzymes that may be useful as herbicides. Nucleic acid fragments can also  
 CC be used as probes for genetically and physically mapping the genes that  
 CC they are a part of, and as markers for traits linked to those genes. Such  
 CC information may be useful in plant breeding in order to develop lines  
 CC with desired phenotypes. 225109 to 225127 represent specifically claimed  
 CC nucleic acids from the present invention and Y42112 to Y42130 represent  
 CC the proteins encoded by them.  
 XX  
 SQ Sequence 603 AA;

Query Match 9.7%; Score 195.5; DB 20; Length 603;  
 Best Local Similarity 26.5%; Pred. No. 5e-10;  
 Matches 61; Conservative 45; Mismatches 105; Indels 19; Gaps 4;

QY 168 QIRAL-AEQANTYHVVLKRHDELFTSATPERLVAMSGQIATAVAGTSRRGTGADDI 226  
 DB 351 EYVALRILVPSPIYMAVGARGCVLVASSPELLIRVSGKILINPLAGTVRRGTKEDEQ 410  
 QY 227 ALGALLASCKNRRIEHOYVASTITRLQDVT--SLKVPAMPSLIKKQVOHLTYPTIGD 284  
 DB 411 MREQQLISDEKQCAEHIMLVLDGRNDYGVKSKPSGVKELIMNIERYSHVMHISSTVSQ 470  
 QY 285 IAAHLSTAYIDRLHPALGCVPREALYIATHEKPRGLFAGPIGFTFADNSGEFVY 344  
 DB 471 LDDHLSWDALRAALPYGVSGAPKVKAMELIDKLEVTIRPYSGYIGYISFDGDMQIAL 530  
 QY 345 GIRSM-----YVNGTGR--ATLFAAGAGIVADSDAQOEYEE 378  
 DB 531 SLRLIVSTAPSHNTMYSKDADTRREWVAHLGAGAGIVADSDDEQRE 580

RESULT 9  
 Y41301  
 ID Y41301 standard; Protein: 616 AA.  
 XX  
 XX Y41301;  
 AC  
 XX 18-JAN-2000 (first entry)  
 DT  
 XX  
 DE Tobacco anthranilate synthase ASA2.  
 DE  
 KW Promoter; tobacco; anthranilate synthase; tolerance; analogue; Trp;  
 KW tryptophan; transgenic plant.  
 XX  
 OS Nicotiana tabacum.  
 XX  
 PN US5965727-A.  
 XX  
 PD 12-OCT-1999.  
 XX  
 PF 31-DEC-1997; 97US-0001826.  
 XX  
 PR 26-JUL-1996; 96US-0025140.  
 PR 25-JUL-1997; 97US-0937739.  
 XX  
 PA (UNIT ) UNIV ILLINOIS FOUND.  
 XX  
 PI Withholm JM, Brotherton JE, Song H;  
 XX  
 DR WPI: 1999-579943/49.  
 DR N-PSDB; 230444.  
 PT Plant tissue culture transformation using promoters and DNA constructs  
 PT -  
 XX  
 PS Example 2: Column 31-34; 49pp; English.  
 XX  
 CC This sequence corresponds to the tobacco (Nicotiana tabacum) anthranilate  
 CC synthase ASA2. The corresponding gene's promoter and DNA constructs  
 CC containing it are useful for imparting, to a plant cell, tolerance to  
 CC an amino acid analogue of Trp, or altering the Trp content in a plant  
 CC by transforming the plant cells with an expression cassette containing  
 CC the ASA2 structural gene. The promoter and construct are useful for the  
 CC production of transgenic plants.  
 XX  
 SQ Sequence 616 AA;

Query Match 9.7%; Score 195; DB 20; Length 616;  
 Best Local Similarity 23.0%; Pred. No. 5.7e-10;  
 Matches 73; Conservative 69; Mismatches 141; Indels 34; Gaps 9;  
 QY 84 EVAVT-----IAADKTFPGSDTSDFTTWLAQFVKQKQNTVTSHTVDEVMIERENTLID 139  
 DB 254 ELIVSVRGIESPLRISPGS---VDFCT--NATGP-----SLTKQNMUSE-eyknavjqake 303  
 QY 140 TLAIQTLAKVVPGRQDTLQISDTLRILAQIRALAEQANTYHVVLKRHDELFTSATPERL 199



```

Db      304 hiaagdtfgivlsqrferrtfadpfeyralrliv--npssymtyiqargcllvasspeil 361
Qy      200 VAMSGGQIATAAAGTSSRRGTGADDIALGBALIASOKNRIEHOYVAVASTTRLODYTT- 258
Db      362 trkxkrrivrrplagtrrgktptedvmlqemgmkldekqraehlmldlgrndvgkyskp 421
Qy      259 -SLKVPAPSLKKNKOYHLYTPTGDIAAHLSYTAIVDRLHPTPALGSGVPREALYYIA 317
Db      422 gsvnveklmverysvnmhlssvsgelldhllcwaalraalpvgvtvsgapkvkamellid 481
Qy      318 THEKTPKGLFACPIGYFTADNSGEFVVGIRSM-----VYNOTQR---ATLF 361
Db      482 qlvartarrgysggfggifsfgmdialalrtmvtlmgarydtmstyrdaaskrgewahlg 541
Qy      362 AGAGIYADSDAQOEYEE 378
Db      542 sgagivadsnpdeeqle 558

```

```

RESULT 10
W93814
ID      W93814 standard; Protein; 606 AA.
AC      W93814;
XX
XX
XX      25-JUN-1999 (first entry)
DT
XX
XX      Rice anthranilate synthase second isozyme alpha-subunit protein.
DE
XX
XX      Anthranilate synthase; alpha-subunit; ASA; rice; isozyme; plant; seed;
KM      maize; wheat; tryptophan content; nutritional value.
XX
XX      Oryza sativa.
OS
XX
XX      WO9911800-A1.
PN
XX
XX      11-MAR-1999.
PD
XX
XX      31-AUG-1998; 98WO-JP03883.
PF
XX
XX      29-AUG-1997; 97JP-0235049.
PR
XX
XX      (HOKU ) HOKKO CHEM IND CO LTD.
PA      (NORO ) JAPAN MIN AGRIC FORESTRY & FISHERIES.
PI      Hasegawa H, Terakawa T, Tozawa Y, Wakasa K;
XX
XX      WPI: 1999-228982/19.
DR      N-PSDB; X23753.
XX
XX      DNA encodes ^a-subunit of first isozyme of rice anthranilate
PT      synthase - used for improving tryptophan production and
PT      nutritional value of crops, e.g. rice, maize or wheat
XX
XX      Claim 3; Page 139-141, 152pp; Japanese.
PS
XX
XX      This invention describes a novel rice anthranilate synthase first
CC      isozyme alpha-subunit. The encoding DNA can be used to produce
CC      transformant plants and seeds, of e.g. rice, maize or wheat, with
CC      enhanced tryptophan content and nutritional value of the crops.
XX
XX      Sequence 606 AA;
SQ

```

Query Match 9.4%; Score 189.5; DB 20; Length 606;  
 Best Local Similarity 26.7%; Pred. No. 1.8e-09;  
 Matches 62; Conservative 45; Mismatches 102; Indels 23; Gaps 6;

```

Qy      168 QIIRAL-AEQANTYHVVLKRHDELFIATPERLIVAMSGGQIATAAAGTSSRRGTGADDI 226
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db      354 evyalrtlvpspymayvgargcvlvasspelltrvrkxklnrplagrtvrrgktekede 413

```

```

Qy      227 ALGBALIASOKNRIEHOYVAVASTTRLODYTT--SLKVPAPSLKKNKOYHLYTPTG 284
Db      414 mteqqlisdekqcaehlmldlgrndvgkyskpsvveklmrleryshnmhlssvsg 473
Qy      285 IAAHLSYTAIVDRLHPTPALGSGVPREALYYIATHEKTPKGLFACPIGYFTADNSGEFV 344
Db      474 lddhlgswdaalraalpvgvtvsgapkvkamellidelevtrrgpysgylglsfd--gdmli 531
Qy      345 GI-----RSMY-VNOTQR---ATLFAGAGIYADSDAQOEYEE 378
Db      532 alalrtlvfstapsnltmnykykdeirrewvahlgagivadsspddegre 583

```

```

RESULT 11
Y42113
ID      Y42113 standard; Protein; 446 AA.
XX
XX      Y42113;
AC
XX
XX      13-DEC-1999 (first entry)
DT
XX
XX      Corn anthranilate synthase alpha subunit portion protein sequence.
DE
XX
XX      Anthranilate synthase alpha subunit; tryptophan biosynthetic enzyme;
KM      chimeric gene; corn; rice; soybean; wheat; tryptophan synthase;
KM      anthranilate synthase beta subunit; herbicide; fungicide; phenotype;
KM      pathogen attack; identification; transgenic plant; ASAS; ASBS; TSAS;
XX      plant breeding.
XX
XX      Zea mays.
OS
XX
XX      WO9949058-A2.
PN
XX
XX      30-SEP-1999.
PD
XX
XX      19-MAR-1999; 99WO-US06046.
PF
XX
XX      26-MAR-1998; 98US-0079386.
PR
XX
XX      (DUPO ) DU PONT DE NEMOURS & CO E I.
PA      Vollmer SJ, Falco SC, Brogile RM, Bryan GT, Cahoon RE;
PI      Rafalski JA;
XX
XX      WPI: 1999-580451/49.
DR      N-PSDB; Z25110.
XX
XX      New isolated tryptophan biosynthetic enzyme nucleic acids, used to
PT      produce plants with altered tryptophan levels and for developing
PT      herbicides or fungicides -
XX
XX      Claim 21; Page 57-59; 83pp; English.
PS
XX
XX      The present invention describes isolated anthranilate synthase alpha-
CC      subunit (ASAS), anthranilate synthase beta-subunit (ASBS), and tryptophan
CC      synthase alpha-subunit (TSAS) nucleic acids, and protein encoded by them,
CC      obtained from corn, rice, soybean and wheat cDNA libraries. The nucleic
CC      acid fragments may be used to create transgenic plants in which the
CC      disclosed ASAS, ASBS or TSAS are present at higher or lower levels than
CC      normal or in cell types or developmental stages in which they are not
CC      normally found. This would have the effect of altering the level of
CC      tryptophan in those cells. Manipulation of the levels of some of the
CC      ASAS will also results in changes in the response to pathogen attack.
CC      Because this pathway is not followed for the production of tryptophan in
CC      higher animals, these enzymes are very good candidates for the discovery
CC      of herbicides and fungicides. The ASAS, ASBS or TSAS can be used as
CC      targets to facilitate design and/or identification of inhibitors of those
CC      enzymes that may be useful as herbicides. Nucleic acid fragments can also
CC      be used as probes for genetically and physically mapping the genes that

```

CC they are a part of, and as markers for traits linked to those genes. Such  
CC information may be useful in plant breeding in order to develop lines  
CC with desired phenotypes. Z25109 to Z25127 represent specifically claimed  
CC nucleic acids from the present invention and Y42112 to Y42130 represent  
CC the proteins encoded by them.

**SQ Sequence 446 AA;**

Query Match	9.3%	Score 186.5;	DB 20;	Length 446;
Best Local Similarity	26.1%;	Pred. No. 2.3e-09;		
Matches	60;	Conservative 45;	Mismatches 106;	Indels 19;
				Gaps 4;

```
Oy 168 QIIRAL-AECANTYHHVLRKHELFISITPERVLAMSGCOIATAVAGTSRGTGDADDI 226
      :| ||| : | : :||| | : :||| ||| |:|
Db 195 eLyralrInpspmaylqargcIlvasspellrtlvqrtilnrplagtlrrgktaeck 254
```

```

255 tleqlllsekqcaehlmvlvgndvykvsbkpsvkveklmteryslhvhhisvtlge 314
db

```

315 lrrdltcwadalraalpvgctvsqapkrivamelldqlvsmrpgpysggfgjlsfrgmdial 374

```
Db      375 alrtlivfptasrfdtmysytdsksrqewahlgagagivadskpddehge 424
```

RESULT	12
W26685	
ID	W26685 standard; Protein; 615 AA.

AC	W26685;
XX	
DT	27-MAR-1998 (first entry)

Maize C28 allele anthranilate synthase alpha subunit

KW Anthranilate synthase; L-tryptophan; tolerance; resistance  
KW transgenic plant; maize; selectable marker.

OS Zea mays cell line C28.

	Key	Location/Qualifiers
FH	Peptide	1..46
FT		
FT		/label= Transit_peptide

PN WO9726366-A1.

PD	24-JUL-1997.
XX	
PF	17-JAN-1997; 97WO-US00983.

PR 19-JAN-1996; 96US-0604789.

PA (DEKA-) DEKALB GENETICS CORP.  
XX  
PI Anderson PC, Chomet PS, Griffor MC, Kriz AL

DR WPI; 1997-385350/35.  
DR N-PSDB; T91102.

PT DNA encoding anthranilate synthase resistant to inhibition by tryptophan - and transformed plants, used e.g. to improve tryptophan levels in plants and as marker for cell selection

PS Claim 4; Fig 7; 87pp; English.

CC This protein sequence comprises a maize anthranilate synthase (AS)  
CC that is resistant to inhibition by free L-tryptophan or its amino  
CC acid analogues. Its amino acid sequence was deduced from a cDNA

CC clone asse T91102) obtained from the 5-methyltryptophan-resistant  
CC maize cell line C28. Compared with the wild-type AS sequence (see  
CC W26868), it contains a single amino acid substitution (Lys for  
CC Met-377) at a site implicated in tryptophan feedback sensitivity of  
CC the enzyme. The C28 allele AS DNA sequence can be used in claimed  
CC methods: (i) to impart tolerance of plants to Trp analogues; (ii)  
CC to alter, particularly increase, Trp content of plants; either to  
CC increase nutritional value or as source of Trp by extraction; (iii)  
CC for production of recombinant AS (used for screening to identify  
CC agents that bind to or inhibit it); and (iv) for selection of  
CC transformed cells. Transgenic plants containing AS can be used for  
CC production of proteins or other compounds, including in vitro  
CC culture of their cells. The trait of resistance to Trp can be  
CC introduced to a wide variety of commercial maize lines.

Sequence 615 AA;

Query Match	9.28;	Score 184.5;	DB 18;	Length 615;
Best Local Similarity	26.08;	Pred. No. 5.5e-09;		
Matches	60;	Conservative 45;	Mismatches 105;	Indels 21;
			Gaps 5;	

```
QY 168 QIIRAL-AEQANTYHVVLKRHDELFISATPERVLAMSGGCIATAVAGTSKRGTGDADDI 2266
    ::|||::|:::|||||::|::|::|
Db 363 eyvalrlivnpspykayvgargcvlvaaspelltrvskgkllnrplagctvrirgktekedq 4222
```

QY 227 ALGEALLASQKNRIEHQYV---ASITTRLDVYTSLSKPRAMPSSLKNQYOHLYPTTG 2833

Db 423 mgegqllsdekqcaehmlvldgrndvqkvskspgsvvkekl-lleryshmhhsstcvsq 4811

QY 284 DIAHLSVFAIVDRLEHPTPALGGVPREALYYIATHEKPPRGLFGPIGITYTADNSEEFV 3433

Db 482 qldhlgswdalraalpvgtyvsgapkvkamelidklevtrrgpysglgisfsgdmgia 5411

```

QY      344 VGIKSM-----YVNOTQRR--ATLFAGAGIADSDAQQEEYE 378
      : : : : : : : : : : : : : : : : : : : : : : : :
Db      542 lslrltvtstapshntmysykddarrrewhanlgagaglvadsspddegre 592

```

RESULT	13
W26686	
ID	W26686 standard; Protein: 615 AA

AC	W26686;
XX	
DT	27-MAR-1998 (first entry)

DE Maize anthranilate synthase alpha subunit.

KW Anthranilate synthase; L-tryptophan; tolerance; resistance  
KW transgenic plant; maize; selectable marker.  
KW

OS Zea mays Inbred line Va26.

	Key	Location/Qualifiers
FH	Peptide	1.46
FT		
FT		/label= Transit_peptide

PN W09726366-A1.

PD	24-JUL-1997.
XX	
PF	17-JAN-1997; 97WO-US00983.

PR 19-JAN-1996; 96US-0604789.

PA (DEKA-) DEKALB GENETICS CORP.

PI Anderson PC, chomet PS, Griffor MC, Kriz AL;

WPI; 1997-385350/35.  
N-PSDB; T91103.



PT nutritional value of crops, e.g. rice, maize or wheat  
XX  
PS Claim 1; Page 128-131; 152pp; Japanese.  
XX

50

YY

Claim 1; Page 128-131; 152pp; Japanese.

YY

CC This invention describes a novel rice anthranilate synthase first  
CC isozyme alpha-subunit. The encoding DNA can be used to produce  
CC transformatant plants and seeds, of e.g. rice, maize or wheat, with  
CC enhanced tryptophan content and nutritional value of the crops.  
CX

XX 0

This invention describes a novel rice anthranilate synthase first isozyme alpha-subunit. The encoding DNA can be used to produce transformant plants and seeds, of e.g., rice, maize or wheat, with enhanced tryptophan content and nutritional value of the crops.

XX 0

This invention describes a novel rice anthranilate synthase first isozyme alpha-subunit. The encoding DNA can be used to produce transformant plants and seeds, of e.g., rice, maize or wheat, with enhanced tryptophan content and nutritional value of the crops.

XX 0

This invention describes a novel rice anthranilate synthase first isozyme alpha-subunit. The encoding DNA can be used to produce transformant plants and seeds, of e.g., rice, maize or wheat, with enhanced tryptophan content and nutritional value of the crops.

XX

Sequence 577 AA;

Query Match  
Best Local

9.08; Score 181.5; DB 20; Length 577;  
25.78; Pred. No. 9.7e-09;  
45; Mismatches 107; T-3-3-10

Best Local Similarity 25.7%; Pred. No. 9.7e-09;  
Matches 59; Conservative 45; Mismatches 107; Indels 19;

Matches	59; Conservative	45; Mismatches	107; Indels	19; Gaps	4;
---------	------------------	----------------	-------------	----------	----

```

QY 168 QIIRAL-AEQANTHVVLKRHDELFIATPERLVAAMSGGQIATVAVGSTRRGDADDI 226
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 326 eyvaltlivpspymlqagaccllvasspelltrvckrtlvmrplagtlrkskaedk 385
QY 227 ALGALLASOKNRIEHQYVVAITRLQDYTT--SLKVPAMPSSLKKNQOVHETPTGD 284
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 386 vlegllisdgqcaehmlvdlgrndvgvkskpsvkvcklmmveryshwmhlsstvtge 445
QY 285 IAAHLSTAIYDRHPPRLALGGVREAAIIYIATHETPRGRLRAGPIGITYTADNCGEYV 344
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 446 lrdldtcwdaLraalprgvtvsagkvramelldqmeqkmtgpyssgfgvsgfmgdmial 505
QY 345 GIRSM-----YVNOQFRR---ATLFAGAGIVADSDAODPEE 378
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 506 alrtlvprgsrfdtmytsycknaergawhlgagagivadsrpdceqne 555

```

5

```

326 eyvriallvnpvpyrnaeyqargcglvvaesspellrtvkerltvnpvplagltirgsksheed 3655
227 AGEALLASOKRIRIEHOYVWASITTRLODYTT--SIYVAMPAPSLILKKOYHOHLTPPTGD 264
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
386 vleglllledgqgaeehmlvdgrndygvkspsvskkekimneyvsnymhslsstvge 4455
285 IAAHLSVYAIRYDRLHPTALGCVPREAALYITATHEKTRPGCLFAGPTGYGTADNNGEEVY 3444
446 lrdldlcwdaIraalpvgvtsqparvzmeilldqmeqkmtgprvsgdfgyasfgydmal 5050

```

2

326 evyrallivnpsymaylgargclivasspelltrvekrctivnrplagllrrgskaedk 385  
227 ALGALLASOKNRIEHQVVASITTRLQDVTT--SLKVPAMPSLLKNQVQHLTPITGD 284

135

```

326 eyrrallivnpsymaylqargcillvaasspelllrvekrliivnrplagllrrrgskaedk 365
227 AGEALLASQKKRIRHOVAVSITTRLODYTT--SLKVPAMPSLKKKOVHLETPITGD 264
      1: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
386 vleqlllsdgkrcaehnlvldgrndvgkvspsvkvkrlmvervshmlsistvtge 445

```

3

```

326 eyrgialvnpwpyrnaylgargcillvaesspelltrvektlrvnrvplogrlgtrrgskbeek 3655
227 ALGELLASOKRIRIHOVAVSITTRLODYTT--SLVYRPMPSILKKOKONHUTPTED 264
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
386 vteqlllsoqgcaehmlvdlgrndvgkvkpsvkvreklmveyrshvmlslstvtge 4455
285 IAAHSVTAIVDRLLHPRTALGSGVPEALYIAAHEKTRPLRGFLGRTGTADNSGEFV 3444

```

5

```

326 eyvriallvnpvpyrnaeyqargcglvvaesspellrtvkerltvnpvplagltirgsksheed 3655
227 ALGELLALSOKRIEHOYVWASITTRLODYTT--SIYVAMPAPSLILKKOYHOHLTPPTGD 264
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
386 vleglllledgqgceehmlvdgrndvgykvskpsvskkekimneyvsnymhslsstvge 4455
285 IAAHLSVYAIYDRLHPTALGCVPREAALYITATHEKTRPGCLFAGPTGTYTADNNGEEV 344
446 lrdldlcwdaIraalpvgvsgarvkrzemeIldqmegkmtgprvsgdfgyasfgydmal 5050

```

50

```

326 eyryaliivnpspymayiqdagcillvasspellirvekrllvmrpiagllirgyskaedk 365
227 ALGELLASQKNRIEHOYVAVSITTRLODVT--SLKVPAMPSLLKKOVONHTPTIGD 284
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
386 vteqlllsdgkgcaeahlmlvdgrndygvkskpsvkvexkmmveryshvmlhsstvtge 445
285 IAAHLSVTAIYDRLHPRTALGVPREAAIYYIAATHEKTPRGLEAFGIVGTADNSGEFVV 344
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
446 lrdldtcwdalaaalpvgvtsgaprvkzamelldgmegkmtgpysgdfgyasffgmdmlal 505
345 GTRSM-----YVNOTOR--ATLEFAGAGIVADSDAODEE 378

```

Db

```

326 eyrrallivnpssymayiqdagccllvasspellirvekrclivmriplagllirgyskaedk 365
227 ALGELLASOKNRLEHOYVAVSITTRLODVT--SLKVPAMPSLLKKOKVOHETPTICD 284
386 vleglllsogkqcaeshlmlvdgrindvgkvskspsvkvcklmmyerishymhlisstvtge 445
285 IAAHLSVTAIYDLRLHPALGCVPREAALYYIAIHEKTRPGLGAPRIGVETADNSGEFVV 344
446 Irdltcdwclaaalpvgvtsgarpvzxamelldgmegkmmrpygsgfygysfgrmdial 505
345 GIKSM-----YVNOTQR--ATLPAAGIATVADSDAODEE 378
506 airtvfpclgsfdclmyziclknaqewallgaacllvadskkddhe 555

```

```
Search completed: November 25, 2000, 02:17:00
Job time: 3222 sec
```

Job time: 3222 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: November 25, 2000, 01:48:48 ; Search time 41.98 Seconds  
(without alignments)  
158.509 Million cell updates/sec

Title: US-08-774-104A-2

Perfect score: 2009

Sequence: 1 MMTYHETALNQSDLOQLXA.....ETGLKPEPMROLKDYHVE 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 16761186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA: \*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/prodata/1/1aa/6\_COMB.pep: \*  
3: /cgn2\_6/prodata/1/1aa/6\_COMB.pep: \*  
4: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep: \*  
5: /cgn2\_6/prodata/1/1aa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2009	100.0	397	2	US-08-774-104A-2	Sequence 2, Appl1
2	210.5	10.5	595	3	US-08-604-789B-3	Sequence 3, Appl1
3	209	10.4	491	2	US-09-001-826-23	Sequence 23, Appl1
4	201.5	10.0	621	3	US-08-604-789B-4	Sequence 4, Appl1
5	195	9.7	616	2	US-09-001-826-5	Sequence 5, Appl1
6	184.5	9.2	604	3	US-08-604-789B-16	Sequence 16, Appl1
7	183.5	9.1	604	3	US-08-604-789B-2	Sequence 2, Appl1
8	96.5	4.8	1346	3	US-09-320-878-4	Sequence 8, Appl1
9	94	4.7	3739	4	PCT-US95-06994-8	Sequence 8, Appl1
10	94	4.7	3739	4	US-09-320-878-2	Sequence 2, Appl1
11	92.5	4.6	416	3	US-08-910-505-2	Sequence 2, Appl1
12	92.5	4.6	416	3	US-08-910-505-4	Sequence 4, Appl1
13	92.5	4.6	1562	3	US-09-320-878-3	Sequence 3, Appl1
14	92.5	4.6	3038	1	US-08-450-332-2	Sequence 2, Appl1
15	92.5	4.6	3038	2	US-08-637-640-2	Sequence 2, Appl1
16	91.5	4.6	4551	3	US-09-320-878-1	Sequence 1, Appl1
17	90.5	4.5	3031	1	US-07-689-008-2	Sequence 2, Appl1
18	90	4.5	636	1	US-08-765-081-5	Sequence 5, Appl1
19	90	4.5	636	1	US-09-098-082-5	Sequence 5, Appl1
20	90	4.5	636	4	PCT-US95-06994-5	Sequence 5, Appl1
21	88	4.4	1167	2	US-08-589-756-2	Sequence 2, Appl1
22	87.5	4.4	487	2	US-08-712-709-8	Sequence 8, Appl1
23	87.5	4.4	487	3	US-09-111-444-8	Sequence 8, Appl1
24	87.5	4.4	1026	1	US-08-194-290-7	Sequence 7, Appl1
25	87	4.3	718	4	PCT-US95-06994-6	Sequence 6, Appl1
26	87	4.3	887	4	US-08-215-709-1	Sequence 1, Appl1
27	86.5	4.3	787	1	US-08-574-763-2	Sequence 2, Appl1
28	85	4.2	800	1	US-08-785-052-4	Sequence 4, Appl1

29	85	4.2	800	2	US-08-913-581-4	Sequence 4, Appl1
30	84	4.2	878	1	US-08-237-919-2	Sequence 2, Appl1
31	84	4.2	878	4	PCT-US95-05518-2	Sequence 2, Appl1
32	83.5	4.2	339	1	US-08-433-854-4	Sequence 4, Appl1
33	83.5	4.2	339	1	US-08-174-745A-4	Sequence 4, Appl1
34	83.5	4.2	339	2	US-08-195-947-4	Sequence 4, Appl1
35	83.5	4.2	339	2	US-08-433-885-4	Sequence 4, Appl1
36	83.5	4.2	339	3	US-08-433-908B-4	Sequence 4, Appl1
37	83.5	4.2	7257	3	US-09-335-409-5	Sequence 5, Appl1
38	83	4.1	1150	2	US-08-589-756-3	Sequence 3, Appl1
39	82.5	4.1	529	1	US-08-548-509-2	Sequence 2, Appl1
40	82.5	4.1	1026	2	US-08-614-377A-7	Sequence 7, Appl1
41	82	4.1	1289	2	US-08-542-003-2	Sequence 2, Appl1
42	82	4.1	1289	2	US-08-322-760A-2	Sequence 2, Appl1
43	81.5	4.1	394	3	US-08-673-814-6	Sequence 6, Appl1
44	81.5	4.1	1164	2	US-08-589-756-1	Sequence 1, Appl1
45	81	4.0	2089	1	US-08-418-893D-23	Sequence 23, Appl1

## ALIGNMENTS

RESULT 1  
US-08-774-104A-2  
Sequence 2, Application US/08774104A  
Patent No. 5919687  
GENERAL INFORMATION:  
APPLICANT: Chatterjee, Subroto  
TITLE OF INVENTION: RECOMBINANT N-SMASES AND NUCLEIC ACIDS  
FILE OF INVENTION: ENCODING SAME  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Dike, Bronstein, Roberts & Cushman, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/774,104A  
FILING DATE: 12/24/96  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Corleiss, Peter F  
REGISTRATION NUMBER: 33,860  
REFERENCE/DOCKET NUMBER: 46906  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 397 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
US-08-774-104A-2  
Query Match 100.0%; Score 2009; DB 2; Length 397;

Best Local Similarity 100.0%; Pred. No. 9,2e-203;  
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MMTYETRALAOSDLOQALYALETEFEAGYFATPADDTLRFGIGATATKKTALOGAVF 60
   |||
Db 1 MMTYETRALAOSDLOQALYALETEFEAGYFATPADDTLRFGIGATATKKTALOGAVF 60
QY 61 GAQSFDEQEPSELMAGFWFVPMVVTIADKITEGSDTVSDFTWLAQFVKOPNTVT 120
   |||
Db 61 GAQSFDEQEPSELMAGFWFVPMVVTIADKITEGSDTVSDFTWLAQFVKOPNTVT 120
QY 121 TSHVTDVDMIRTEMLIDTLAIDOTLAKVFGROQTQLSPTLRILAQIIRALAEQANTY 180
   |||
Db 121 TSHVTDVDMIRTEMLIDTLAIDOTLAKVFGROQTQLSPTLRILAQIIRALAEQANTY 180
QY 181 HVLKRBHDELFSATPERLYAAMSGGIATAAAGTSRRGTGADADIALGELLASOKNRI 240
   |||
Db 181 HVLKRBHDELFSATPERLYAAMSGGIATAAAGTSRRGTGADADIALGELLASOKNRI 240
QY 241 EHQYVASTITRLODYTTSLKVPAMPSSLKKNQOVHLYPTIGDIAHLSTVATVDRLLP 300
   |||
Db 241 EHQYVASTITRLODYTTSLKVPAMPSSLKKNQOVHLYPTIGDIAHLSTVATVDRLLP 300
QY 301 TPALGVPREALYIATHEKTPRGLFAGPIGYFTADNSGEFVVGIRSMYVNOTORRATL 360
   |||
Db 301 TPALGVPREALYIATHEKTPRGLFAGPIGYFTADNSGEFVVGIRSMYVNOTORRATL 360
QY 361 FAGAGIVADSDAQOEYEETGLKFEPRKOLLKDYNHVE 397
   |||
Db 361 FAGAGIVADSDAQOEYEETGLKFEPRKOLLKDYNHVE 397

```

## RESULT 2

US-08-604-789B-3  
Sequence 3, Application US/08604789B  
Patent No. 6118047

GENERAL INFORMATION:  
APPLICANT: Anderson, P.C.

Chomet, P.S.  
Griffor, M.C.  
Kilz, A.L.

TITLE OF INVENTION: ANTHRANILATE SYNTHASE GENE  
AND METHOD OF USE THEREOF FOR CONFERRING TRYPTOPHAN  
OVERPRODUCTION

NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A.  
STREET: P.O. Box 2938  
CITY: Minneapolis

STATE: MN  
COUNTRY: USA  
ZIP: 55402

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/604,789B  
FILING DATE: 19-Jan-1996

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: <Unknown>

FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:

NAME: Woessner, Warren D.  
REGISTRATION NUMBER: 30,440

REFERENCE/DOCKET NUMBER: 950.026US1  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (612) 373-6903  
TELEFAX: (612) 339-3061

TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 595 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-08-604-789B-3

Query Match 10.5%; Score 210.5; DB 3; Length 595;

Best Local Similarity 26.0%; Pred. No. 1.7e-13;  
Matches 76; Conservative 57; Mismatches 124; Indels 35; Gaps 8;

```

QY 110 QFVPKQNTVTTSHVTDVDMIRTEMLIDTLAIDOTLAKVVF-----GROQTQLSDPL 164
   |||
Db 294 QFVPSLDN-----SNVTC-----EYKAVYK--AKEHILAGDIFQIYLSQRFERRTADPF 343
QY 165 RLAQIIRALAEQANTYHVLKRBHDELFSATPERLYAAMSGGIATAAAGTSRRGTGAD 224
   |||
Db 344 EYVRAIRV--NPSPYMGYLQARGCILVASSPEILTKVKONKIVNRPPLAGTSKRGKNEVE 401
QY 225 DIALGELLASOKNRIEHQYVASTITRLODYTT--SLKVPAMPSSLKKNQOVHLYPTIT 282
   |||
Db 402 DKRLKELLENEKQCAHEHMLVDLGRNDGKTVKGSVVEKLMNIERSHVHHISSTVT 461
QY 283 GDIAHLSTVATVDRLLPPTPALGVPREALYIATHEKTPRGLFAGPIGYFTADNSGEF 342
   |||
Db 462 GELQDLDLTCDVYRALPLPGTSGAPKVKAMELIDLEPTRRKPGYSGFGGVSFTGDDMI 521
QY 343 VVGIRSM-----YVNOTORR--ATLFAGAGIVADSDAQOEYE 378
   |||
Db 522 ALSRTIVFPACQYNTMYSKDANKRRREVAVLQAGAGVADSDPDQDCE 573

```

## RESULT 3

US-09-001-826-23  
Sequence 23, Application US/09001826A  
Patent No. 5965727

GENERAL INFORMATION:  
APPLICANT: SONG, HEE-SOOK

APPLICANT: BROUGHTON, JEFFREY E.  
APPLICANT: WIDHOLM, JACK M.

TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE  
FILE REFERENCE: UI001.C1

CURRENT APPLICATION NUMBER: US/09/001,826A  
CURRENT FILING DATE: 1997-12-31

EARLIER APPLICATION NUMBER: 08/937,739; 60/025,140  
EARLIER FILING DATE: 1997-07-25; 1996-07-26

NUMBER OF SEQ ID NOS: 31  
SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text ;

SEQ ID NO 23  
LENGTH: 491

TYPE: PRT  
ORGANISM: Nicotiana tabacum

US-09-001-826-23

Query Match 10.4%; Score 209; DB 2; Length 491;

Best Local Similarity 26.4%; Pred. No. 1.8e-13;  
Matches 77; Conservative 56; Mismatches 123; Indels 36; Gaps 9;

```

QY 110 QFVPKQNTVTTSHVTDVDMIRTEMLIDTLAIDOTLAKVVF-----GROQTQLSDPL 164
   |||
Db 191 QFVPSLDN-----SNVTC-----EYKAVYK--AKEHILAGDIFQIYLSQRFERRTADPF 240
QY 165 RLAQIIRALAEQANTYHVLKRBHDELFSATPERLYAAMSGGIATAAAGTSRRGTGAD 224
   |||
Db 241 EYVRAIRV--NPSPYMGYLQARGCILVASSPEILTKVKONKIVNRPPLAGTSKRGKNEVE 298
QY 225 DIALGELLASOKNRIEHQYVASTITRLODYTT--SLKVPAMPSSLKKNQOVHLYPTIT 282
   |||
Db 299 DKRL-ELLENEKQSAHEHMLVDLGRNDGKTVKGSVVEKLMNIERSHVHHISSTVT 357

```



```

1      Kriz, A.L.
2      TITLE OF INVENTION: ANTHRANILATE SYNTHASE GENE
3      AND METHOD OF USE THEREOF FOR CONFERRING TRIPTOPHAN
4      OVERPRODUCTION
5
6      NUMBER OF SEQUENCES: 16
7
8      CORRESPONDENCE ADDRESS:
9      ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A.
10     STREET: P.O. Box 2938
11     CITY: Minneapolis
12     STATE: MN
13     COUNTRY: USA
14     ZIP: 55402
15
16     COMPUTER READABLE FORM:
17     MEDIUM TYPE: Diskette
18
19     COMPUTER: IBM Compatible
20
21     OPERATING SYSTEM: DOS
22
23     SOFTWARE: FastSeq Version 2.0
24
25     CURRENT APPLICATION DATA:
26     APPLICATION NUMBER: US/08/604,789B
27     FILING DATE: 19-Jan-1996
28
29     PRIOR APPLICATION DATA:
30     APPLICATION NUMBER: <Unknown>
31     FILING DATE: <Unknown>
32
33     ATTORNEY/AGENT INFORMATION:
34     NAME: Woessner, Warren D.
35     REGISTRATION NUMBER: 30,440
36
37     REFERENCE/DOCKET NUMBER: 950.026U1
38
39     TELECOMMUNICATION INFORMATION:
40     TELEPHONE: (612) 373-6903
41     TELEFAX: (612) 339-3061
42     TELEX: <Unknown>
43
44     INFORMATION FOR SEQ ID NO: 16:
45     SEQUENCE CHARACTERISTICS:
46     LENGTH: 604 amino acids
47     TYPE: amino acid
48     STRANDEDNESS: single
49     TOPOLOGY: linear
50     MOLECULE TYPE: protein
51     SEQUENCE DESCRIPTION: SEQ ID NO: 16:
52
53     US-08-604-789B-16
54
55     Query Match          9 28;   Score 184.5;   DB 3;   Length 604;
56     Best Local Similarity 26.08;   Pred. No. 9.3e-11;
57     Matches 60;   Conservative 45;   Mismatches 105;   Indels 21;   Gaps 5;
58
59     QY 168 QIIRAL-AEQNTAVVILKRDELFIISATPEERIVAMSGGQIATAAAGTSRRGTGADDI 226
60           ::|||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
61     DB 352 EYVALRLRVNSPKAIVQARGCVLVASSPEILIRTSKGIINRPLAGIVRRKRTKEQD 411
62
63     QY 227 ALGALLASQKRIEHQYV--ASITRLQDVTTSLIKVPAMSLKKNQOVHLYPTIG 283
64           ::|:|  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
65     DB 412 MQEQQLLSDEKQAEHIMLVDLGRNDYGVSKPGGSVYKVEKL-IIRYSVHMIISSTVSG 470
66
67     QY 284 DIAHLSVTAIVDLHPTRPALGVPKRALYYTATHEKTRGLFAGPIGTFYADNSGCEV 343
68           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
69     DB 471 QLDHILQSWDALRALAPVGTVSGAPKYKAMELIDKLEVTIRRGPSGGLGISFDGMQIA 530
70
71     QY 344 VGIRSM-----YVNTQIR--ATLFPAGAGIADSDAQGEYE 378
72           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
73     DB 531 LSLTIYFSTAPSHNTMYSTKDDRRKREWYAHILQAGIYADSSPDDEORE 581
74
75     RESULT 7
76     US-08-604-789B-2
77     Sequence 2, Application US/08604789B
78     Patent No. 6118047
79
80     GENERAL INFORMATION:
81     APPLICANT: Anderson, P.C.
82             Chomet, P.S.
83             Grilfor, M.C.
84             Kriz, A.L.
85
86     TITLE OF INVENTION: ANTHRANILATE SYNTHASE GENE

```

```

AND METHOD OF USE THEREOF FOR CONFERRING TRYPTOPHAN
OVERPRODUCTION

NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A.
STREET: P. O. Box 2938
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/604,789B
FILING DATE: 19-Jan-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <unknown>
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Woessner, Warren D.
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 950.026US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (612) 373-6503
TELEFAX: (612) 339-3061
TELEX: <unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 604 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-604-789B-2

Query Match 9.1%; Score 183.5; DB 3; Length 604;
Best local similarity 26.0%; Pred. No. 1.2e-10;
Matches 60; Conservative 45; Mismatches 105; Indels 21; Gaps 5;

QY 168 QIIRAL-AEQANTYHVVLKRHDELISATPERLIAMSGQIATAAVAGTSRRGTGADDI 226
::||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 352 EYVRALRIVNSPYMAVYQARGCVLVAASPEILIRVSKGIINRPLAGTVRRGKTEREDQ 411
::||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 227 ALGERLLASQCNRIEHQYV---ASTIRLQDVTTSLKVPAMPSLKNKOYOLYTPITG 283
::|:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 412 MQEOQLLSDKEQCAEHMLVDLGRNDYKSKPGSVYKVEKL-IIEEYSHVMIISSIVSG 470
::|:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 284 DIANHLSTATVLDRLHPRLPGVPREALYYLTATHEKTPRGLFAGPIGYFTADNNGSEFV 343
::|:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 471 QLDHLSWDALRALPLVGTSGAPKYKAMELIDKLEVYTRRGPGYSGGSGISFDGMQIA 530
::|:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 344 VGIRSM-----YVNTQQR---ATTFAGAGIYADSDAQGEYEE 378
::|:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 531 LSLRTVETASPTNMTWYSYKADNRREWRVAHLQAGAGIYADSDSPDEQRE 581
::|:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
US-09-320-878-4
: Sequence 4, Application US/09320878A
: Patent No. 6117659
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanie C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
: APPLICANT: TANG, Li
: TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
: FILE REFERENCE: 300622002120

```



```

: CURRENT APPLICATION NUMBER: US/09/320, 878A
: CURRENT FILING DATE: 1999-05-27
: EARLIER APPLICATION NUMBER: CIP OF 09/141, 908
: EARLIER FILING DATE: 1998-08-28
: EARLIER APPLICATION NUMBER: CIP OF 09/073, 538
: EARLIER FILING DATE: 1998-05-06
: EARLIER APPLICATION NUMBER: CIP OF 08/846, 247
: EARLIER FILING DATE: 1997-04-30
: EARLIER APPLICATION NUMBER: 60/119, 139
: EARLIER FILING DATE: 1999-02-08
: EARLIER APPLICATION NUMBER: 60/100, 880
: EARLIER FILING DATE: 1998-09-22
: EARLIER APPLICATION NUMBER: 60/087, 080
: EARLIER FILING DATE: 1998-05-28
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: Patent Ver. 2.0
: SEQ ID NO 4
: LENGTH: 1346
: TYPE: PRT
: ORGANISM: Streptomyces venezuelae
US-09-320-878-4
```

```

Query Match          4.8%; Score 96.5; DB 3; Length 1346;
Best Local Similarity 24.2%; Pred. No. 0.62;
Matches 73; Conservative 30; Mismatches 118; Indels 81; Gaps 15;
```

```

QY 3 TYHETRALAOSDLOO-LYALETTEFGAYFPYPPADDTLRFSGIATATKTAQALOGAVEG 61
D 523 TAMEHRAVAVGDSREALDALRME-GIVRGTVDP-----GRVAFFPQGGTQWAGM 575
QY 62 AQSDEDEYPOSELMAGWFPEYVWVTIAADKITFGSDTVSDFTTWLAQFPKQPNVYTT 121
D 576 AELLDS-----PEFAAAMECE-----TALSPYDWSLEAVVROAPSAPT 616
QY 122 SHVDEVMIERTENLIDTLAIDQTLAKV-----VGRQO-----TLQLS 161
D 617 ---LDRDVYQPV-----TFVWVSLAKVQHHGHTPEAVIGHSGOGLAAVYVAGALTLD 668
QY 162 D-----TLRLAQIIRALAEQANTYHVVLKRHDEFISATPERLYAMSGGQIATAVAGTS 216
D 669 DAARVTVLRKSIAMHLAGKGMISLASE-----EATRQRIENLHG--LSIAAVNGPT 720
QY 217 RRGTDGADDTLALGALLASQKNRI-----EHQIVASITTRLODYVTTSL--KYPA 264
D 721 ATTVSG-DPTQIQELAOACEADGIRARIIPYDVASHSAHVETIENELADVLAGLSPQTPQ 779
QY 265 MP 266
D 780 VP 781
```

```

RESULT 9
PCT-US95-06994-8
: Sequence 8, Application PC/TUS9506994
: GENERAL INFORMATION:
: APPLICANT: Children's Hospital & Medical Center
: APPLICANT: University of Washington
: APPLICANT: Washington State University Research Foundation
: APPLICANT: TARR, PHILIP I
: APPLICANT: BILGE, SIMA S
: APPLICANT: BESSER, THOMAS E
: APPLICANT: VART JR, JAMES C
: TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPILLC
: STREET: SUITE 2800, 1420 FIFTH AVENUE
: CITY: SEATTLE
: STATE: WASHINGTON
: COUNTRY: USA
: ZIP: WA 98101
: COMPUTER READABLE FORM:
```

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/06994
: FILING DATE: 07-JUN-95
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/265, 714
: FILING DATE: 24-JUN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: BRODERICK, THOMAS F
: REGISTRATION NUMBER: 31,332
: REFERENCE/DOCKET NUMBER: CHOR-18591
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 682 8100
: TELEFAX: (206) 224 0779
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 703 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: DESCRIPTION: E. coli O157:H7 adhesin amino acid sequence,
: DESCRIPTION: wherein "Xaa" residues represent gaps
: DESCRIPTION: introduced to facilitate best alignment with
: HYPOTHETICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli O157:H7
: STRAIN: 86-24 NAHR
PCT-US95-06994-8
```

```

Query Match          4.7%; Score 94; DB 4; Length 703;
Best Local Similarity 22.5%; Pred. No. 0.39;
Matches 76; Conservative 37; Mismatches 125; Indels 100; Gaps 16;
```

```

QY 63 QSFDEQYPOSELMAGWFPEYVWVTIAADKITFGSDTVSDFTTWLAQFPKQPNVYTT 122
D 265 QRYDRDQQLSLNGY-----DRLRYERNKTSAGYDGTFTGWTXSY----- 308
QY 123 HVTDEVMIERTENLIDTLAIDQTLAKVFGQQTQLSDTLRLAQIIRALAEQANTYHV 182
D 309 -----LNMNE--TENK-----GRELVRSYLKR-----DKWGLAGOPRELKES-----NL 345
QY 183 VLKRHDEFISATPERLYAMSGGQ-----IATAVAGTSRRGTD-----GADDLALG 229
D 346 IL--NSLLTPTPLGESHLYVGGEFQSSMKDGVVLASTGETFEROKSKWSVFAEDEMHLT 402
QY 230 EALLASQKNRIHQ-----YVVASITTRLODYVTTSLKVPAMPISLKNKQVQ 275
D 403 DALATAGSREYHEHQFGHFSPRAYIWDVADAMTLKGGVTVGKAPRMQGLHNG----- 458
QY 276 HLYTPTGDIAHLSVTALVDRLHPTPALG--VPREALYYITTHEKTPRGFLPAGPIG 332
D 459 -----ISG-----VSGOGKTNLGNPDLPEBSVSYEAGVYVDNPAGLMANVYG 502
QY 333 YFTADNNGEFYVGI-----RSMYVNGQTOR--ATLFAG 363
D 503 FMT-DFSNKKIXYSINDNTNSIVNSGKARLHGVERAG 539
```

```

RESULT 10
US-09-320-878-2
: Sequence 2, Application US/09320878A
: Patent No. 611/659
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanie C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
```

```

; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3739
; TYPE: PRF
; ORGANISM: Streptomyces venezuelae
US-09-320-878-2

```

Query Match 4.7%; Score 94; DB 3; Length 3739;

Best Local Similarity 22.4%; Pred. No. 6.1; Mismatches 143; Indels 128; Gaps 20;

```

Matches 89; Conservative 38;
QY 6 ETRALA-----QSDLOQLYALLETTEFGAYFATPADTLRFGIGAIATAKTAQALOGAVFG 61
DB 2033 EHRAVALGTGODDLAALALAPESGLVRG-----VASGVRAVAFVFGGTQWAGMG 2082
QY 62 AQSFDQEQPQSLMAGFMFVPEVMTIADKTTFGSDTVSDFTTMAQVPRQPTVT 121
DB 2083 AELLDVSK-----EFAAMAMECEA-----AAPYDVDSLVAHQAPGAPT 2123
QY 122 SHVTDEVDWIERTENLIDTLAIDOTLAKV-----VFGROQ-----TLQLS 161
DB 2124 LERVNDVGVY-----TFAVVSLAKWQHNGVTPQAVVGHSGGEIAAAYVAGALSLD 2175
QY 162 D-----TLRLAQITRALAQDANTYHVVLKRHDELFTSATPERLVAMSGGOIATAVAGTS 216
DB 2176 DAARVYTLKRSKSGIHALACGCMISLASE-----AAVERLAGFDG--LSVAAYVNGPT 2227
QY 217 RRGTDGADDIAGELLASOKNRI-----EHQYVASTITRLQDYTTSL--KVPA 264
DB 2228 ATIVVSG-DPTQIOELAQACEADGVRRARIIPVDYASHSAHETTESLADVLGLSPQTPQ 2286
QY 265 MP--SLKKN-----KOVQHL--YTPITGDIANLSVTAIVD-RLHP--TPA 303
DB 2287 VPFESTLEGAWITERPALDGCYWRNLRHVGFAVETLADTDEGTFHFEVSAHPVLITMA 2346
QY 304 LCGVPREALYYIATHEKTPRGILFACPIGYFTADNSGE 341
DB 2347 L-----PEYVTGL-----GTLRDRNGGQ 2364

```

```

RESULT 11
US-08-910-505-2
; Sequence 2, Application US/08910505A
; Patent No. 6107071
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie
; APPLICANT: Reichard, Raymond
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith

```

```

; APPLICANT: Lonetto, Michael
; APPLICANT: Warren, Patrick
; TITLE OF INVENTION: NOVEL HISTIDINOL DEHYDROG ENASE
; FILE REFERENCE: P50549-02
; CURRENT APPLICATION NUMBER: US/08/910,505A
; CURRENT FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 416
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-08-910-505-2

```

Query Match 4.6%; Score 92.5; DB 3; Length 416;

Best Local Similarity 20.1%; Pred. No. 0.23; Mismatches 154; Indels 99; Gaps 20;

```

Matches 80; Conservative 66;
QY 5 HETRALAQSDLOQLYALLETTEFGAYFATPADTLRFGIGAIATAKTAQALOGAVFGAQS 64
DB 35 HGDKALKMYNLTPDHTDTHLEISHQIKAAFDTL-----DETKQALQOSYERIK 86
QY 65 FDE-----QEPQS-ELMAGFMFVPEVMTIADKTTFGSDTVSDFTTMAQ----- 110
DB 87 YQESIKOTNOQLSESVCEIYHPLBSGIVYVPGKASYSYVL--MTATLAQAVGENTI 144
QY 111 --FVPRQPNVT-----TSNVTDEVDWIERTENLIDTLAIDOTLAKVFGRODTLSDPT 163
DB 145 VVVTPEQPNVSGSEVLAACYITG-----VNOVFQVGAQSIATLTVG--TETIPKVD- 194
QY 164 LRLAQITRALAQDANTYHVVLKRHDELFTSATPERLVAMSGGOIATAVAGTSRSG--T 220
DB 195 -----KIVPGNQGFAAATKX--LF-----GQVGDQIAGPTEIALIID 231
QY 221 DGAD-DIALGELLASOKNRIEHQYVASTITRLQDYTTSLKVPAMSL--LKNQV- 274
DB 232 DTADLDIAIVYDFAQAEHDELARTYVIGEDAQLKDELSRI-AKALPNVDRYDIVSKSIA 290
QY 275 -QH-LVTPITGDIANLSV-----AYDRLHPTRPALGVPREALYYIATHEKTPRGILF 327
DB 291 NOHYLLHASNFDKACHVMNTIAPENHASIQVNPQPYTEKKYVAGLF----- 337
QY 328 AGPIGYFTADNSGEFVVGIRSMYVNCORRATFEAGAGI 366
DB 338 ---IGHSPVIGDYVAG--PSHVLPRNRTARRFTNGLSV 371

```

```

RESULT 12
US-08-910-505-4
; Sequence 4, Application US/08910505A
; Patent No. 6107071
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie
; APPLICANT: Reichard, Raymond
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith
; APPLICANT: Lonetto, Michael
; APPLICANT: Warren, Patrick
; TITLE OF INVENTION: NOVEL HISTIDINOL DEHYDROG ENASE
; FILE REFERENCE: P50549-02
; CURRENT APPLICATION NUMBER: US/08/910,505A
; CURRENT FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 416
; TYPE: PRF

```

; ORGANISM: Staphylococcus aureus  
US-08-910-505-4

US-08-910-505-4

Query Match	4.68;	Score 92.5;	DB 3;	Length 416;
Best Local Similarity	20.18;	Pred. No. 0.23;		
Matches	80;	Conservative	66;	Mismatches 154;
				Indels 99;
				Gaps 20;

[illegible]

RESULT 13  
US-09-320-878-3  
; Sequence 3, Application US/09320878A

```

? GENERAL INFORMATION:
? APPLICANT: ASHLEY, Gary
? APPLICANT: BETLACH, Melanie C.
? APPLICANT: BETLACH, Mary C.
? APPLICANT: MCDANIEL, Robert
? APPLICANT: TANG, Li
? TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLKETIDE SYNTHASE
? FILE REFERENCE: 300622002120
? CURRENT APPLICATION NUMBER: US/09/320,878A
? CURRENT FILING DATE: 1999-05-27
? EARLIER APPLICATION NUMBER: CIP OF 09/141,908
? EARLIER FILING DATE: 1998-08-28
? EARLIER APPLICATION NUMBER: CIP OF 09/073,538
? EARLIER FILING DATE: 1998-05-06
? EARLIER APPLICATION NUMBER: CIP OF 08/846,247
? EARLIER FILING DATE: 1997-04-30
? EARLIER APPLICATION NUMBER: 60/119,139
? EARLIER FILING DATE: 1999-02-08
? EARLIER APPLICATION NUMBER: 60/100,880
? EARLIER FILING DATE: 1998-09-22
? EARLIER APPLICATION NUMBER: 60/087,080
? EARLIER FILING DATE: 1998-05-28
? NUMBER OF SEQ ID NOS: 34
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 3
? LENGTH: 1562
? TYPE: PRT
? ORGANISM: Streptomyces venezuelae
? US-09-320-878-3

```

Query Match 4.68; Score 92.5; DB 3; Length 1562;

Best Local Similarity 20.38; Pred. No. 2.1;  
Matches 71; Conservative 46; Mismatches 120; Indels 113; Gaps 13

```

0Y      3 TYHERBALAQSDLOQ-LVAALETTEFGAYFAPPPADDDTLRFIGIAIATKTAQNALOGANFG 61
      1 | | | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      526 TAMEHRAAVALGDSREALDRLAMP-EGLVGRGTSS-----VGRAEYFPGOGTOWAGM 578
0Y      62 AOSFDEQEQPOSELMAGFMEFPEVWVTTIAADKIFFGSDTSDFTYWLQFOPKOPNTVT 122
      1 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      579 AELDDSS-----PEFASABEE-----TALSRYDMSLEAVROBPAP 611
      1 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
0Y      122 SHVTEVDVMIERTENILITLAIIDQTLAKVFGEGROQTQLSDTLRLAQITRALAEQANTYH 183
      1 | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
Db      620 ---LDRVVQPV-----TFAMVSLAK----- 633
0Y      182 VYLKRHDELFTSAPPERLVAMSGGOIATAAAG-----TSR-----RGTDG, 222
      1 | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
Db      640 -VMOHH-----GIFQAVAVGHSGEGIAAAYVALTLDDAARVYTLRKSIAHLAKGKG 693
0Y      223 ADDIATGERALLASQKNRIE-----HOYVASITTRLODYTSLKAVAMBSLKNK 273
      1 | | | | : : : | | | | | | | | | | : : : | | | | | | | | | | | | | | |
Db      694 MISTALDEBAALVKRLSDPDGLSVAAVNGPRTAVVSGDPTQIEELARTCEADGVARI--- 755
0Y      273 QVQNHLYPTGDIANHLSTVALVDRLHPRLPALGVCPREALUYUATHEKT 322
      1 | : | | | | | | | | | | | | | | : | : | : | : | : | : | : | : | : |
Db      751 -----LPDYASHSRQVEITIEELAEVLGLAGLAPHPHFSTLEGT 792

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

ORIGINAL SOURCE:  
ORGANISM: TPKS Protein  
US-08-450-332-2

Query Match 4.6%; Score 92.5; DB 1; Length 3038;  
Best Local Similarity 23.5%; Pred. No. 6.2;  
Matches 89; Conservative 36; Mismatches 123; Indels 131; Gaps 20;

QY 12 QSDLOQLYALLETTERGAFATPADTLRFGIGAIATAKT-----AQALOG---AVF-GA 62  
DB 74 QDGLGAFDAFFNIQAGE--AESMDPOHRLLETVYEAVTNAGMRIODLOGSTAYVGV 131  
QY 63 QSFDEQEPQSELMAGFWFVPEVMTIADKITEGSDTVSDFTTWLAQFVKPQNTVTS 122  
DB 132 MTHDYETVSTRDLES-----IPYTSATGVA--VSVASNRISYFFDW----- 170  
QY 123 HTDEVDMIERENLIDTLAIDOTLAKVVGROQTLOLSDTLRLAQIIRALAEQANTYHV 182  
DB 171 -----HGPMSTIDT--ACSSSLVAVHLAYQ-----LRTGQSSMAIAGAN----- 209  
QY 183 VLKRHDELFTSATPERLVAMSGQIATAVAGTSRRCTGADADIALGEALLASOKNRIEH 242  
DB 210 -----LILGPMTEFVLES--KLSMLSPSGSRMRMDAGADGVARGEAVCS----- 250  
QY 243 QYVVASITTRLOD-----VTSILKVP---AMPSILKKNQOVHLYTPIT 282  
DB 251 -VVLKTLSQLARQDRTIECVIRETVGNQDRTTGITMPNHSADALIKATYAO----- 302  
QY 283 GDIAHLSTVAIVDR-----LHPTPALGVPREALYYIAT---HEKTPRG-----L 326  
DB 303 ---AGLDTIKAKDRCOFFFAHGTGPAGDPQEAFA--IATAFFGHQVARSQGNERAPL 356  
QY 327 FAGPIGFTADNSGEFVVG 345  
DB 357 FVG-----SAKTVVG 366

RESULT 15  
US-08-637-640-2

Sequence 2, Application US/08637640

Patent No. 5849541

GENERAL INFORMATION:

APPLICANT: VINCI, VICTOR A.

APPLICANT: CONDER, MICHAEL J.

APPLICANT: MCADA, PHYLLIS C.

APPLICANT: REEVES, CHRISTOPHER D.

APPLICANT: DAVIS, CHARLES R.

APPLICANT: HENDRICKSON, LEE E.

APPLICANT: RAMBOSEK, JOHN

TITLE OF INVENTION: DNA ENCODING TRIOL POLYKETIDE SYNTHASE

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: CHRISTINE E. CARTY

STREET: P. O. BOX 2000; 126 E. LINCOLN AVENUE

CITY: RAHWAY

STATE: NJ

COUNTRY: USA

ZIP: 07065

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/637,640

FILING DATE: 23-AUG-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/148,132

FILING DATE: 01-NOV-1993

ATTORNEY/AGENT INFORMATION:

NAME: CARTY, CHRISTINE E.

REGISTRATION NUMBER: 36,099  
REFERENCE/DOCKET NUMBER: 19076  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-594-6734  
TELEFAX: 908-594-4720  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3038 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: TPKS Protein  
US-08-637-640-2

Query Match 4.6%; Score 92.5; DB 2; Length 3038;  
Best Local Similarity 23.5%; Pred. No. 6.2;  
Matches 89; Conservative 36; Mismatches 123; Indels 131; Gaps 20;

QY 12 QSDLOQLYALLETTERGAFATPADTLRFGIGAIATAKT-----AQALOG---AVF-GA 62  
DB 74 QDGLGAFDAFFNIQAGE--AESMDPOHRLLETVYEAVTNAGMRIODLOGSTAYVGV 131  
QY 63 QSFDEQEPQSELMAGFWFVPEVMTIADKITEGSDTVSDFTTWLAQFVKPQNTVTS 122  
DB 132 MTHDYETVSTRDLES-----IPYTSATGVA--VSVASNRISYFFDW----- 170  
QY 123 HTDEVDMIERENLIDTLAIDOTLAKVVGROQTLOLSDTLRLAQIIRALAEQANTYHV 182  
DB 171 -----HGPMSTIDT--ACSSSLVAVHLAYQ-----LRTGQSSMAIAGAN----- 209  
QY 183 VLKRHDELFTSATPERLVAMSGQIATAVAGTSRRCTGADADIALGEALLASOKNRIEH 242  
DB 210 -----LILGPMTEFVLES--KLSMLSPSGSRMRMDAGADGVARGEAVCS----- 250  
QY 243 QYVVASITTRLOD-----VTSILKVP---AMPSILKKNQOVHLYTPIT 282  
DB 251 -VVLKTLSQLARQDRTIECVIRETVGNQDRTTGITMPNHSADALIKATYAO----- 302  
QY 283 GDIAHLSTVAIVDR-----LHPTPALGVPREALYYIAT---HEKTPRG-----L 326  
DB 303 ---AGLDTIKAKDRCOFFFAHGTGPAGDPQEAFA--IATAFFGHQVARSQGNERAPL 356  
QY 327 FAGPIGFTADNSGEFVVG 345  
DB 357 FVG-----SAKTVVG 366

Search completed: November 25, 2000, 02:17:58  
Job time: 1750 sec



**THIS PAGE BLANK (USPTO)**

Mon Dec 4 10:44:32 2000

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2000, 01:55:53 ; Search time 51.06 Seconds  
(without alignments)  
493.414 Million cell updates/sec

Title: US-08-774-104a-2

Perfect score: 2009

Sequence: 1 MMTYHETRALAQSDQLQLYA.....ETGLKFEPRLQKDYNHVE 397

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

PIR 65: \*  
1: pirl: \*  
2: pirl2: \*  
3: pirl3: \*  
4: pirl4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	496.5	24.7	471	1	A69657
2	453	22.5	476	1	S58229
3	383	19.1	474	1	S58229
4	354.5	17.6	372	2	H70595
5	344	17.1	391	1	SYECIX
6	328.5	16.4	398	1	C69615
7	326.5	16.3	396	1	A40365
8	317.5	15.8	356	2	G64997
9	289	14.4	421	2	JC5323
10	278.5	13.9	523	2	A42301
11	278	13.8	430	2	E64059
12	272.5	13.6	456	2	S35124
13	257	12.8	516	2	G70556
14	253	12.6	511	2	T36306
15	252.5	12.6	433	2	T43924
16	251.5	12.5	434	2	A75163
17	249.5	12.4	470	2	A37854
18	246.5	12.3	494	2	E70352
19	246	12.2	464	2	C69088
20	245	12.2	461	2	D72414
21	245	12.2	515	1	JNBS1
22	244.5	12.2	512	2	JH0098
23	244	12.1	462	2	A40362
24	236	11.7	529	2	T45254
25	234.5	11.7	317	2	B72489
26	232	11.5	474	2	B64434
27	227	11.3	474	2	T35072
28	223.5	11.1	416	2	H81376
29	220	11.0	497	2	S11891

30	219	10.9	508	2	S74537	anthranilate synth
31	218.5	10.9	489	2	T40974	anthranilate synth
32	217.5	10.8	454	2	A31132	p-aminobenzoate sy
33	217	10.8	494	2	JX0065	anthranilate synth
34	217	10.8	508	2	S19266	anthranilate synth
35	216.5	10.8	434	2	T17434	anthranilate synth
36	215.5	10.7	638	2	F75547	anthranilate synth
37	215	10.7	465	2	S75655	anthranilate synth
38	215	10.7	503	2	T46852	anthranilate synth
39	214.5	10.7	473	2	E75353	anthranilate synth
40	213.5	10.6	411	2	B69450	anthranilate synth
41	213.5	10.6	507	1	NNBY1	anthranilate synth
42	211.5	10.5	491	2	E81132	anthranilate synth
43	210.5	10.5	364	2	T43181	probable anthranil
44	210.5	10.5	491	2	G81892	probable anthranil
45	210.5	10.5	595	2	J01684	anthranilate synth

## ALIGNMENTS

RESULT 1  
A69657  
probable isochorismate synthase (EC 5.4.99.6) menaquinone-specific menF - Bacillus su  
C:Species: Bacillus subtilis  
C>Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 16-Jun-2000  
C:Accession: A69657; S27507; S27508; I39883; T46638; T46639  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gal  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holstappel, S.; Hosono, S.; Hullo, M  
Koelter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masudaj, S.; Mau  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudeg, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portere  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl  
A:Authors: Schleich, S.; Schroeter, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Se  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Toynoni, A.; Tosato, V.; Uchiya  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; M01D:98044033  
A:Accession: A69657  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-471 <KUN>  
A:Cross-references: GB:299119; GB:AL009126; NID:g2635411; PIDN:CAB15061.1; PID:g26355  
A:Experimental source: strain 168  
R:Rowland, B.; Hill, K.; Mueller, J.; Driscoll, J.; Taber, H.  
Submitted to the EMBL Data Library, October 1991  
A:Description: Organization of an operon involved in menaquinone biosynthesis in Bac  
A:Reference number: S27507  
A:Accession: S27507  
A:Molecule type: DNA  
A:Residues: 1-11, 'K', 13-110, 'WDLYSEDFLLTLAKKVNHTSRKGLSLRLC' <ROW1>  
A:Cross-references: EMBL:M74538; NID:g1185287  
A:Molecule type: DNA  
A:Residues: 151-471 <ROW2>  
A:Cross-references: EMBL:M74538; NID:g1185287  
R:Miller, P.  
J. Bacteriol. 170, 2742-2748, 1988  
A:Title: Transcriptional regulation of a promoter in the men gene cluster of Bacillus  
A:Reference number: I39883; M01D:88227858  
A:Accession: I39883  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-11, 'K', 13-32 <ML1>  
A:Cross-references: GB:M21320; NID:g143178; PIDN:AAA22594.1; PID:g551715  
R:Driscoll, J.R.; Taber, H.W.  
J. Bacteriol. 174, 5063-5071, 1992  
A:Title: Sequence organization and regulation of the Bacillus subtilis menBE operon.

A:Reference number: A42715; MUID:92332443  
A:Accession: T46638  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-11, 'K', '13-110, '134-135, 'LCYSE', '141-142, 'LL', 'T', '317', 'ANKEVHN', '325, 'T', '327', '328', '329', '330', '331', '332', '333', '334', '335', '336', '337', '338', '339', '340', '341', '342', '343', '344', '345', '346', '347', '348', '349', '350', '351', '352', '353', '354', '355', '356', '357', '358', '359', '360', '361', '362', '363', '364', '365', '366', '367', '368', '369', '370', '371', '372', '373', '374', '375', '376', '377', '378', '379', '380', '381', '382', '383', '384', '385', '386', '387', '388', '389', '390', '391', '392', '393', '394', '395', '396', '397', '398', '399', '400', '401', '402', '403', '404', '405', '406', '407', '408', '409', '410', '411', '412', '413', '414', '415', '416', '417', '418', '419', '420', '421', '422', '423', '424', '425', '426', '427', '428', '429', '430', '431', '432', '433', '434', '435', '436', '437', '438', '439', '440', '441', '442', '443', '444', '445', '446', '447', '448', '449', '450', '451', '452', '453', '454', '455', '456', '457', '458', '459', '460', '461', '462', '463', '464', '465', '466', '467', '468', '469', '470', '471', '472', '473', '474', '475', '476', '477', '478', '479', '480', '481', '482', '483', '484', '485', '486', '487', '488', '489', '490', '491', '492', '493', '494', '495', '496', '497', '498', '499', '500', '501', '502', '503', '504', '505', '506', '507', '508', '509', '510', '511', '512', '513', '514', '515', '516', '517', '518', '519', '520', '521', '522', '523', '524', '525', '526', '527', '528', '529', '530', '531', '532', '533', '534', '535', '536', '537', '538', '539', '540', '541', '542', '543', '544', '545', '546', '547', '548', '549', '550', '551', '552', '553', '554', '555', '556', '557', '558', '559', '560', '561', '562', '563', '564', '565', '566', '567', '568', '569', '570', '571', '572', '573', '574', '575', '576', '577', '578', '579', '580', '581', '582', '583', '584', '585', '586', '587', '588', '589', '590', '591', '592', '593', '594', '595', '596', '597', '598', '599', '600', '601', '602', '603', '604', '605', '606', '607', '608', '609', '610', '611', '612', '613', '614', '615', '616', '617', '618', '619', '620', '621', '622', '623', '624', '625', '626', '627', '628', '629', '630', '631', '632', '633', '634', '635', '636', '637', '638', '639', '640', '641', '642', '643', '644', '645', '646', '647', '648', '649', '650', '651', '652', '653', '654', '655', '656', '657', '658', '659', '660', '661', '662', '663', '664', '665', '666', '667', '668', '669', '670', '671', '672', '673', '674', '675', '676', '677', '678', '679', '680', '681', '682', '683', '684', '685', '686', '687', '688', '689', '690', '691', '692', '693', '694', '695', '696', '697', '698', '699', '700', '701', '702', '703', '704', '705', '706', '707', '708', '709', '710', '711', '712', '713', '714', '715', '716', '717', '718', '719', '720', '721', '722', '723', '724', '725', '726', '727', '728', '729', '730', '731', '732', '733', '734', '735', '736', '737', '738', '739', '740', '741', '742', '743', '744', '745', '746', '747', '748', '749', '750', '751', '752', '753', '754', '755', '756', '757', '758', '759', '760', '761', '762', '763', '764', '765', '766', '767', '768', '769', '770', '771', '772', '773', '774', '775', '776', '777', '778', '779', '780', '781', '782', '783', '784', '785', '786', '787', '788', '789', '790', '791', '792', '793', '794', '795', '796', '797', '798', '799', '800', '801', '802', '803', '804', '805', '806', '807', '808', '809', '810', '811', '812', '813', '814', '815', '816', '817', '818', '819', '820', '821', '822', '823', '824', '825', '826', '827', '828', '829', '830', '831', '832', '833', '834', '835', '836', '837', '838', '839', '840', '841', '842', '843', '844', '845', '846', '847', '848', '849', '850', '851', '852', '853', '854', '855', '856', '857', '858', '859', '860', '861', '862', '863', '864', '865', '866', '867', '868', '869', '870', '871', '872', '873', '874', '875', '876', '877', '878', '879', '880', '881', '882', '883', '884', '885', '886', '887', '888', '889', '890', '891', '892', '893', '894', '895', '896', '897', '898', '899', '900', '901', '902', '903', '904', '905', '906', '907', '908', '909', '910', '911', '912', '913', '914', '915', '916', '917', '918', '919', '920', '921', '922', '923', '924', '925', '926', '927', '928', '929', '930', '931', '932', '933', '934', '935', '936', '937', '938', '939', '940', '941', '942', '943', '944', '945', '946', '947', '948', '949', '950', '951', '952', '953', '954', '955', '956', '957', '958', '959', '960', '961', '962', '963', '964', '965', '966', '967', '968', '969', '970', '971', '972', '973', '974', '975', '976', '977', '978', '979', '980', '981', '982', '983', '984', '985', '986', '987', '988', '989', '990', '991', '992', '993', '994', '995', '996', '997', '998', '999', '1000', '1001', '1002', '1003', '1004', '1005', '1006', '1007', '1008', '1009', '1010', '1011', '1012', '1013', '1014', '1015', '1016', '1017', '1018', '1019', '1020', '1021', '1022', '1023', '1024', '1025', '1026', '1027', '1028', '1029', '1030', '1031', '1032', '1033', '1034', '1035', '1036', '1037', '1038', '1039', '1040', '1041', '1042', '1043', '1044', '1045', '1046', '1047', '1048', '1049', '1050', '1051', '1052', '1053', '1054', '1055', '1056', '1057', '1058', '1059', '1060', '1061', '1062', '1063', '1064', '1065', '1066', '1067', '1068', '1069', '1070', '1071', '1072', '1073', '1074', '1075', '1076', '1077', '1078', '1079', '1080', '1081', '1082', '1083', '1084', '1085', '1086', '1087', '1088', '1089', '1090', '1091', '1092', '1093', '1094', '1095', '1096', '1097', '109

Query Match	24.78;	Score 496.5;	DB 1;	Length 471;
Best Local Similarity	30.28;	Pred. No. 2.4e-30;		
Matches 127;	Conservative 77;	Mismatches 161;	Indels 55;	Gaps 10;

```

0Y 19 YAALETFEGGAVFNRPADDTLRFGLGALATPKTQ----- 53
      | | | | | | | | | | | | | | | | | | | | | |
Db 46 YGAKKTYNNRFFMSDBREBELTVGLGAEAVQTQKKSERXREVEDEMRKTAHITYE 105
      | | | | | | | | | | | | | | | | | | | | | |
0Y 54 -----ALOGAVFGAOSFDEQEPQSE-----LMAGFWEVEVMTYADK-----17 95
      | | | | | | | | | | | | | | | | | | | | | |
Db 106 EKKLQHSASVGVULRGGSFDCBERGSGQMDHFSBGDFVRLMLTMADEPRLTVNKMVS 165
      | | | | | | | | | | | | | | | | | | | | | |
0Y 96 FGSD---TVSDPTWMLAQ-VP---KQRP---LVTTSHYTDVMDYTERENLIDPLAIDQTL 147
      | | | | | | | | | | | | | | | | | | | | | |
Db 166 GGEBAEAVLEGLKFAFEMFWRDKQEDQAVIAAEELDKDWMKALETATSOIK-EKQY 224
      | | | | | | | | | | | | | | | | | | | | | |
0Y 148 AKVFGROOTQLSDTLRLAQIIPALA-EQANTYHVULKRHDELFIATPBRIVAMSGQ 206
      | | | | | | | | | | | | | | | | | | | | | |
Db 225 DKVULARELLTFEPGPIQIEVULFTLDDQOTSIVFAIEBGCKTFGASBERLIRKRDGT 284
      | | | | | | | | | | | | | | | | | | | | | |
0Y 207 IATAVACTSRRCGTGDGADIALGELLASQKNRIEHOYVVAISTTRLADYTSCLKPAMP 266
      | | | | | | | | | | | | | | | | | | | | | |
Db 285 VMSGSLGASIKRGVNEEDRIRIGLELLDENKENLLEHDIYVGMIHNAVSSCSEYKPDGP 344
      | | | | | | | | | | | | | | | | | | | | | |
0Y 267 SLKKNKOVQHLTYPIITGDIAAHLSVTAIVDRLHPALGGVPREALYVITATHEKTPRGL 326
      | | | | | | | | | | | | | | | | | | | | | |
Db 345 VLYKTSVQHLTFPIYVGLRESASIFDLIEKLTPTPALGSSPOEKADVIRETEPMSRGW 404
      | | | | | | | | | | | | | | | | | | | | | |
0Y 327 FAGRPCTYTDANSGSEFVYVGTSMVNVNQTORATLFAAGACIYAOSDAQOEVEFGKFEPM 386
      | | | | | | | | | | | | | | | | | | | | | |
Db 405 YAAPIGWMDSDQNEEFANRSLGIEGSTAR--TFACCGIVEOSEPTSEKEEVOIKIKPM 462
      | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 2

S58229

pcha protein - *Pseudomonas aeruginosa*

C:Species: *Pseudomonas aeruginosa*

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S60203; S58329

R:Serino, L.; Reilmann, C.; Baur, H.; Beyeler, M.; Vlisca, P.; Haas, D.

Mol. Gen. Genet. 249, 217-228, 1995

A:Title: Structural genes for salicylate biosynthesis from chorismate in *Pseudomonas aeruginosa*

A:Reference number: S60202; MUID:96086939

A:Accession: S60203

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-476 <SE2>

A:Cross-references: EMBL:X82644; NID:q1628425; PIDN:CAAS7969.1; PID:q928780

A>Note: this protein is involved in salicylate biosynthesis

A:Superfamily: isochorismate synthase

Query Match	22.5%;	Score 453;	DB 1;	Length 476;
Best Local Similarity	42.4%;	Pred. No. 5.2e-27;		
Matches 103;	Conservative 38;	Mismatches 98;	Indels 4;	Gaps 3;

QY	149	KVYGRROTLLOLSTLRLAQIIRRL- AEQANTYIVULKRNDELFI STPFRPYLVMSSGOI	207
Db	221	KVLARTDARPLGD- TEFWOYIEHLRIQHADAOL FACRKNACPLGSPERLVRIRGEA	279
QY	208	ATAAVAGTSRRGTGDADDIALGEALLA SOKNRLEHQYVVASITTRLODVTLS KVPAMES	267
Db	280	LTHALAGTIARGGDAQEDARLGALL DLSADKRNHEQLVEAIRTALEP FEFSEVLEIPDAPC	339
QY	268	LKKNOVOHLETPITGDIAHLSTAYAL VDRLHPRLAGVYPREALLYIXIT HETKTRGLE	327
Db	340	LKRLARVOHLETPILRARLADAGVIL RLRLQALHPPRPAVGGT PRSAALDYIQHOGMDGW	399
QY	338	AGPIGYFTADNSGEEVYVNGIRSMY NOTQRATLFPAGAIYADDAQOEY EETGLKEEPMR	387
Db	400	AAPICMLDGEENGDELVALRFSALL- -TPRGXILEFAGCGLVDSPEAHYE RETCLKLSAMR	457
QY	388	QLL 390	
Db	458	EAL 460	

RESULT 3  
575568  
isochlorismate synthase (EC 5.4.99.6) - *Synechocystis* sp. (strain PCC 6803)  
N:Alternate names: protein slr0817  
C:Species: *Synechocystis* sp.

C; Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C; Accession: S75568  
R; Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,  
O., K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasui,  
DNA Res. 3, 109-126, 1996

A:Title: sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
 5.  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S75568  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-474 <KAN>  
 A:Cross-references: EMBL:D90911; GB:AB001339; NID:q1653083; PIDN:BAAL8129.1; PID:q16550  
 A:Note: the nucleotide sequence was submitted to the EMBL Data library, June 1996  
 C:Genetics:  
 A:Gene: entC  
 C:Superfamily: isochorismate synthase  
 C:Keywords: intramolecular transferase; isomerase

Query Match	19.1%;	Score 383;	DB 1;	Length 474;
Best Local Similarity	31.5%;	Pred. No. 1.2e-21;		
Matches 112;	Conservative 66;	Mismatches 150;	Indels 28;	Gaps 9;

[illegible]



Db 421 SEGNAEFVIGIRSLAL--SRNRRLVYAGAGIVAGSDPLKEVAIEELKQTLMSRLI 474

RESULT 4

H70595 Probable entc protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: H70595

R:COLE, S.T.; Brosch, R.; Parhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 359, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Bartell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:9825987

A:Accession: H70595

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-372 <COL>

A:Cross-references: GB:295120; GB:AL123456; NID:g3261739; PIDN:CAB08301.1; PID:g2072670

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: entC

C:Superfamily: Isochorismate synthase

Query Match 17.6%; Score 354.5; DB 2; Length 372;

Best Local Similarity 36.6%; Pred. No. 1.2e-19;

Matches 102; Conservative 35; Mismatches 129; Indels 13; Gaps 6;

Db 85 PKRVAAALPPADYDLRIGRANDLLAIFGDPKRVLLAAVQTLADAPLRLRLV 144

QY 175 EQANTYHVVV-----KRHDE---LFTSATPERLVAMSGOIATAAAGTSRGTGADDI 226

Db 145 VADPTAVGYLVDTLSAGNDTGAALVAGSPELLVARSNGNVKCFPGSAPRADAPRLDA 204

QY 227 ALGELLASQKNRIHQYVVAITRLQDVTSLKVPAMSLKKNQVQHLVPTIGDIA 286

Db 205 ANAAALASSAKNRHEHDLVDTMKVLALEPLCEPLTTPAQPOLNNTAAVWMLCTAITRL- 263

QY 287 AHSVTAI--VDRLHPPLAGVPREAALYYIATHEKTPRGRLAGPGYGTADNSGSEVY 344

Db 264 RNTSTTALDALALHPPPAGVPTKATLLELE- GDRGFTAGAVGCDGSGDGMVY 322

QY 345 GRSMTVNOTORRATLFPAGAGIVADSDAQOEYEETGLKF 383

Db 323 SIRCAQLSADRRAALAHAGGIVAESDPDELETTTKF 361

RESULT 5

SYECIK Isochorismate synthase (EC 5.4.99.6) [validated] - Escherichia coli

N:Alternate names: isochorismate synthetase

C:Species: Escherichia coli

C:Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 20-Apr-2000

C:Accession: J0497; S04323; I73520; G64792

R:Ozenberger, B.A.; Brickman, T.J.; McIntosh, M.A.

J. Bacteriol. 171, 775-783, 1989

A:Title: Nucleotide sequence of Escherichia coli isochorismate synthetase gene entc and

A:Molecule type: DNA

A:Residues: 1-304, 'TA', 307-391 <EIK>

A:Cross-references: GB:M36700; EMBL:X12670; NID:g145839; PIDN:AAA18491.1; PID:g145840

R:Brickman, T.J.; Ozenberger, B.A.; McIntosh, M.A.

J. Mol. Biol. 212, 669-682, 1990

A:Title: Regulation of divergent transcription from the iron-responsive feb-entc pro

A:Reference number: 156426; MUID:90230305

A:Accession: 173520

A:Status: preliminary; translated from GB/EMBL/DDB

A:Molecule type: DNA

A:Residues: 1-33 <RES>

A:Cross-references: EMBL:X53274; NID:g48747; PIDN:CAA37371.1; PID:g48749

R:Battner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: G64792

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-391 <BLAT>

A:Cross-references: GB:AE000165; GB:U00096; NID:g1786808; PIDN:AAC73694.1; PID:g17868

A:Experimental source: strain K-12, substrain MG1655

C:Comment: This enzyme catalyzes the isomerization of chorismate to isochorismate, an

of the entc gene is regulated by iron and possibly by the product of the enta gene, 2

C:Genetics:

A:Gene: entC; febP; menP

A:Map position: 14 min

C:Function:

A:Description: EC 5.4.99.6 [validated; MUID:89123153]

A:Pathway: enterobactin biosynthesis; menaquinone biosynthesis

C:Superfamily: isochorismate synthase

C:Keywords: enterobactin biosynthesis; intramolecular transferase; isomerase

Query Match 17.1%; Score 344; DB 1; Length 391;

Best Local Similarity 28.6%; Pred. No. 8.3e-19;

Matches 113; Conservative 65; Mismatches 153; Indels 64; Gaps 12;

Db 19 YAALETTFGAYFATPA-----DITLRFGIGATATKTAOLAGAVE-GAOSFDEQEP 71

QY 72 QSELMAQFWVPEVMTIADKITFGSDIYSDFTTMAQVPPKQPNVTTSHTVDEVDWI 131

Db 86 SS-----LYTPE-----SW--QSFSSRQKQASARRPFR----- 111

QY 132 ERTENLIDTLAI--DQTLAKVVFGRQTLQLS--DTRLAQIIRALAEQANTYHVVTKR- 186

Db 112 SQSLNVERQALPEQTTFEQMVARAAALITATPOVDKVLKSLIDITTDAAIDSGVLERL 171

QY 187 -----HDEL-----FISATPERLVAMSGOIATAAAGTSRGTGADDIAGEA 231

Db 172 IAQNPVSYNFHVPLADGVLGASPELLLRKDERFSSIPLAGSARQPEVDLREAGNR 231

QY 232 LMSQKRRIHQYVVAITRLQDVTSLKVPAMSLKKNQVQHLVPTIGDIAHLV 291

Db 232 LMSQKRRIHQYVVAITRLQDVTSLKVPAMSLKKNQVQHLVPTIGDIAHLV 291

QY 292 TAIYDRILHPPLAGVPREAALYYIATHEKTPRGRLAGPGYGTADNSGSEVY 351

Db 292 LTIACLIHPPLAGVPREAALYYIATHEKTPRGRLAGPGYGTADNSGSEVY 351

QY 352 NOTORRATLFPAGAGIVADSDAQOEYEETGLKFEP 386

Db 352 RENQVR--LFGAGIVPASSPLGEMRETGVKLSM 384

RESULT 6

C69615 Isochorismate synthase dhbc - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000

C:Accession: C69615  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bertelle, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Caprano, V.; Carter, N.M.; Chen, C.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabel, C.; Ferrari, E. *Nature* 390, 249-256, 1997  
A:Authors: Foulger, D.R.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallere, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Iech, J.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Meneel, Y. M.; Ogawa, K.; Oglawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Pottebell, Rieger, M.; Rivolta, S.; Kocho, E.; Kocho, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, T.; A:Authors: Schliebl, C.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serotina, akuchil, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipet, A.; Yamamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshida, A:Authors: Yoshikawa, H.F.; Zunsstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:96044033  
A:Accession: C69615  
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residue length: 1-398 <RUN>  
A:Cross-references: GB:299120; GB:AL009126; NID:g2635613; PIDN:CAB15189.1; PID:g2635696  
A:Genetic source: strain 168  
C:Genetics:  
A:Gene: dhdc  
C:Superfamily: Isochorismate synthase

Query Match	16.4%	Score 328.5	DB 1	Length 398
Best Local Similarity	28.5%	Pred. No. 1.3e-17		
Matches 111; Conservative	64;	Mismatches 163;	Indels 51;	Gaps 12

```

0Y 28 GAVE-ATPADDTLRFSGICAI-----TAKTAQLOGAVFGAOSQDEQYQSE 74
Db 22 GAFFLASHRHVLANKICEIYVPADGONOMETLSRIRIABLRQAKOSQS----- 71
QY 75 LMAGFWEPEVWVITIAADKITFGSDTVSFTTWLA--OF--VPKOPVTVTSH---VYDE 127
Db 72 -----RPLVYGAVPFDQVKAARLVPEEVRMSGRLQFDHEEKQQAQHYYHIKVPREP 124
QY 128 VDMIRTERTENLIDTLAIDOTLAKVFGROQTLOLSPTLRKAOITRALAQAN---TYHAVL 184
Db 125 EDYKNGVEQGLARIR-DOITLSKIVLSRSLHLTSPERPIDTELLRHIAOHNSIGYFEADV 183
QY 185 KRHDE-----LEISATPERIVAMSGOJATAAVACTSRGTDGADDIALGEMILLASQNR 239
Db 184 SSOETSPRRITLLGASPELVLSRMGTQVANSPLASGRPSRNDPYEDQRRAAELSSPARDL 243
QY 240 IEHQYVVASITTRLODDVYTSLKVPAMPSPILKNKQVOHQHYTPITGDIANHLSVTA--IYDR 297
Db 244 HEHAVVAADVAALARPFCRTLEVPREKPELKIETMMHLSVYKGL-SDPSTYALIELAA 302
QY 298 LHPTRALCGVPREALYYIATHEKTPRGLFAGPIGYFTADNSGEFVAGIRSMYVNOQOR 357
Db 303 LHPTRAVGCTPDLAREAILSTIEPDRGFFGTGMVCMCDAGGEWIVITR-CAEERS 360
QY 358 ATLFAGAGIVADSDAQOEYEETGLKFEPM 386
Db 361 LRLVAGAGVAGSKPEDELOETSAKFRM 389

```

RESULT 7  
A40365  
siderophore biosynthetic protein amoA - *Aeromonas hydrophila*  
C:Species: *Aeromonas hydrophila*  
C:Date: 10-Sep-1999 #sequenceRevision 10-Sep-1999 #textChange 10-Sep-1999  
A:Accession: A40365  
R:Barhouth, S.; Payne, S.M.; Arceneaux, J.E.L.; Byers, B.R.  
J. Bacteriol. 173, 5121-5128, 1991  
A:Title: Cloning, mutagenesis, and nucleotide sequence of a siderophore biosynthetic gene  
A:Reference number: A40365, MUID:91317731  
A:Accession: A40365  
A:Status: Preliminary

A: Molecule type: DNA  
A: Residues: 1-396 <BAR>  
A: Cross-references: GB:M63339  
A: Note: the authors translated the codon GAG for residue 393 as Gly  
C: Superfamily: isochozimate synthase

Query Match	16.3%	Score 326.5	DB 1	Length 396
Best Local Similarity	35.28%	Pred. No. 1.8e-17		
Matches 87	Conservative 43	Mismatches 112	Indels 5	Gaps 4

[illegible]

QY	384	EPMRQLL	390
		:	
Db	382	GTMLKAL	388

RESULT 8  
G64997  
isochorismate synthase (EC 5.4.99.6) - *Escherichia coli*  
C:Species: *Escherichia coli*  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Oct-1999  
A:Accession: G64997; S68696; S61203  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.: Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12 .  
A:Reference number: A64720; MUID:97426617  
A:Accession: G64997  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-356 <BIAT>  
A:Cross-references: GB:AA000316; GB:U00096; NID:g1788594; PIDN:AACT5325.1; PID:g178866  
A:Experimental source: Strain K-12, Substrain MC155  
R:Mueller, R.; Dahm, C.; Schulte, G.; Leistner, E.  
FEBS Lett. 378, 131-134, 1996  
A:Title: An isochorismate hydroxymutase isogene in *Escherichia coli*.  
A:Reference number: S68696; MUID:96140724  
A:Accession: S68696  
A:Molecule type: DNA  
A:Residues: 'MLTTRRE', 30-93, 'E', 95-356 <MUE>  
A:Cross-references: EMBL:Z50849; NID:g1050906; PID:g1050907  
A:Experimental source: Strain MC4100  
C:Genetics:  
A:Gene: menP  
C:Keywords: Intramolecular transferase; isomerase

```

Query Match          15.8%  Score 317.51  DB 2;  Length 356;
Best Local Similarity 30.3%  Pred. No. 7.6e17;
Matches 106;  Conservative 51;  Mismatches 162;  Indels 31;  Gaps 10.

QY 59 VFGAQSFDSEQETPOSELMAGFEWFEVEMVTIADITFGSDIVSDFT-----TWLA 109
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db 19 IWMGLAFDPDSQ-----GNLLRPLEMRRCGGKATLRITLFSSSLQHDHAIQKKEFLA 70
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::

QY 110 QFVPKQP-----NTVTTSHTVDEVDWTERENLIDTLAIDQTLAKVYFGROOTIQLSDFT 163
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::

```

Db 71 TLVSIKPLRGHLTTTREQHMPDKGTQJLIELATKRIA-EGELDKVYLVARLARDLHPAS 129

QY 164 LRLAQIIRALAE-QANTYHVVLKKNDE-LTISATPELYAMSGQIATAAVGSTRGTD 221

Db 130 VNAAAMAAASRRLLINCYHFMFARDGENAFSGSPERLMRRRDKALTELTAGTAVANNPD 189

QY 222 GADDIALGELLASOKRIEHQYVVASITTRLODVTSLLKVPMSLLKKQVQHLYTP1 281

Db 190 DKQAQQLGEMLMADKKQREMLVVEIDICRLADTDTLTV-LPQVYLIRKKQVHLRCT 248

QY 282 TGDIAAHLSTVTAIVDRLLHPPTALGGVREALALYIATHEKTPRGFLFAGPIGYFTADNCGE 341

Db 249 WTSINKRADVICTL-HQLOPFAAIVAGLPRDLARQEIARHNPFTREWAGASAGYLSLQGS-E 306

QY 342 FVVGIRSMYVNOJORRATLRFAGAIIVADSDAOGEYETGLKFERMQDLK 391

Db 307 FCVSRIRSKSIGSNVVR-LVAGAGIIVGSDPEDEMQDIDKKAGLRLLTQ 354

RESULT 9  
JC5323

n:anthranilate synthase [EC 4.1.3.27] component I [validated] - *Sulfolobus solfataricus*  
 N:Alternate names: anthranilate synthase alpha chain  
 C:Species: *Sulfolobus solfataricus*  
 C:Date: 15-May-1997 #sequence\_revision 18-Jul-1997 #text\_change 17-Mar-2000  
 C:Accession: J05323; A06035  
 R:Tufano, M.L.; Tosco, A.; Marino, G.; Sanna, G.  
 Biochem. Biophys. Res. Commun. 230, 306-310, 1997  
 A:Title: Expression of *Sulfolobus solfataricus* *trpE* and *trpG* genes in *E.coli*.  
 A:Reference number: J05323; MUID:97168965  
 A:Accession: J05323

A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-421 <run>  
R:Turtino, M.L.; Searano, G.; Marino, G.; Samia, G.; Cubellis, M.V.  
J. Bacteriol. 175, 299-302, 1993  
A:Title: Tyryptophan biosynthesis genes tpeEG in the thermoacidophilic archaeobacterium S  
A:Reference number: A40635; MUID:93106970

A:Contents: MT-4  
A:Accession: A40635  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-421 <Tt2>  
A:Cross-references: GB:M98048; NID:g152936; PIDN:AAA7379.1; PID:g152937  
A:Note: sequence extracted from NCBI backbone (NCBIN:121234, NCBIP:121235)  
C:Genetics:  
A:Gene: trpE

C:Complex: hetrootetramer; two component I chains (PIR:JC5323), two component II chains  
C:Function: <ANT>  
A:Description: EC 4.1.3.27 [validated; MUID:97168965]  
A:Pathway: tryptophan biosynthesis  
A:Note: first step  
C:Function: <COM1>  
A:Description: EC 4.1.3.27 [validated; MUID:99380543]  
A:Note: magnesium required  
C:Superfamily: anthranilate synthase component I  
Keywords: carbon-carbon lyase; oxo-acid-lyase; tryptophan biosynthesis

Query Match	14.48	Score 289	DB 2	Length 421
Best Local Similarity	24.98	Pred. No. 1.5e-14		
Matches 95	Conservative 72	Mismatches 163	Indels 52	Gaps 11

```

0Y      36 DDTRFGAIGAIATAKTAQ---ALOGAFGAQSPD-----EOEYFQSEIMAG 78
      ||      ||      ||      ||      ||      ||      ||      ||
Db      56 DDPVILNGYLKDKLADIPGLFKGGMIGIYSDAVRFMEKIRDKRAADWYAE---- 111
      ||      ||      ||      ||      ||      ||      ||      ||
0Y      79 FMEVY-----EWVTTAADKITGSDVTSPTTWLAQFAPKQPNVTYTSHVDE-V 128
      ||      ||      ||      ||      ||      ||      ||      ||
Db      112 -FTTPDNIITIDHNEGKYYVADLSSVGG--CGIGEFKVSFDESINKKSYRIVASESI 168
      ||      ||      ||      ||      ||      ||      ||      ||
0Y      129 DWIERTEMLIDTLAIDOTLAKVYFGR-QQTQTSDDLTLAQTIRALAEQANTYHVYLKRH 187
      ::      ::      ::      ::      ::      ::      ::      ::

```

[illegible]

RESULT	10
A42301	

anthranilate synthase (EC 4.1.3.27) alpha chain - Haloterrax volcanii  
C.Species: Haloterrax volcanii  
C.Date: 10-Jul-1992 #sequence\_revision 18-Sep-1992 #text\_change 22-Jun-1999  
C.Accession: A42301  
R.Iam, W.L.; Logan, S.M.; Doolittle, W.F.  
J. Bacteriol. 174, 1694-1697, 1992  
A.Title: Genes for tryptophan biosynthesis in the halophilic archaeobacterium Haloterrax volcanii  
A.Reference number: A42301; MUID:92165748  
A.Accession: A42301  
A.Status: preliminary

A: Molecule type: DNA  
A: Residues: 1-523 <LAM>  
A: Cross-references: GB:M83788; NID:q149036; PIDN:AAA73177.1; PID:q149039  
A: Note: the authors failed to give the translation for ACC in residue 239 as shown in C: Superfamily: anthranilate synthase component I  
C: Keywords: carbon-carbon lyase; oxo-acid lyase

Query Match	13.9%	Score 278.5	DB 2	Length 523
Best Local Similarity	28.9%	Pred. No. 1.3e-13		
Matches	79	Conservative	115	Indels 35
				Gaps 5

[illegible]

## RESULT 11

E64059  
 Probable isochorismate synthase (EC 5.4.99.6) - Haemophilus influenzae (strain Rd kw24)  
 C.Species: Haemophilus influenzae  
 C.Date: 18-Aug-1995 #sequence  
 R.Revision: 18-Aug-1995 #text  
 C.Date: 16-Jul-1999  
 C.Accession: E64059  
 R.Fletcher, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirnesh, E.F.; Kerlavage  
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodde, A.; Kelley, J.M.; Weidman  
 ; D.M.; Brando, R.C.; Fine, L.D.; Fritchman, J.L.; Geoghagen, N.S.







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2000, 02:18:04 ; Search time 47.14 Seconds  
(without alignments)  
269.057 Million cell updates/sec

Title: US-08-774-104A-2

Perfect score: 1 MMTHETRALAQSDLOQLYA.....ETGLKFEPMROLKDYNHVE 397

Sequence: 2009

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	496.5	24.7	471	1	MENF_BACSU
2	453	22.5	476	1	PCNA_PSEAE
3	344	17.1	391	1	ENHC_ECOLI
4	328.5	16.4	398	1	DHBC_BACSU
5	326.5	16.3	396	1	AMOA_AERHY
6	317.5	15.8	431	1	MENF_ECOLI
7	308.5	15.4	395	1	VIBC_VIBCH
8	289	14.4	421	1	TRPE_SUISO
9	278.5	13.9	523	1	TRPE_HALVO
10	278	13.8	430	1	MENF_HAFIN
11	272.5	13.6	456	1	TRPE_LACLA
12	257	12.8	516	1	TRPE_MYCTU
13	253	12.6	511	1	TRPE_STRKO
14	252.5	12.6	433	1	TRPE_PYRKO
15	249.5	12.4	470	1	PABB_ADSU
16	246.5	12.3	494	1	TRPE_AODAE
17	246	12.2	464	1	TRPE_METH
18	245	12.2	461	1	TRPE_THEMA
19	244.5	12.2	515	1	TRPE_BACSU
20	244.5	12.2	513	1	TRPE_BACPU
21	244	12.1	462	1	TRPE_METTM
22	236	11.7	529	1	TRPE_MYCLE
23	234.5	11.7	438	1	TRPE_AERPE
24	232	11.5	474	1	TRPE_METJA
25	226	11.2	531	1	TRPE_ARTGO
26	220	11.0	497	1	TRPE_ACICA
27	219	10.9	508	1	TRPE_BACST
28	219	10.9	508	1	TRPE_SYNY3
29	217.5	10.8	454	1	PABB_SALTY
30	217	10.8	454	1	TRPE_CLOTM
31	217	10.8	454	1	TRPE_BACCA
32	215	10.7	485	1	TRF2_SYNY3
33	214	10.7	500	1	TRPE_RHOSH

34	213.5	10.6	411	1	TRPE_ARCFU	028669 archaeoglob
35	213.5	10.6	506	1	TRPE_YEAST	P00899 saccharomyc
36	211.5	10.5	491	1	TRPE_NEIMB	P56995 neisseria m
37	210.5	10.5	491	1	TRPE_NEIMA	Q9xaz0 neisseria m
38	210.5	10.5	491	1	TRPE_NEIMC	Q9s358 neisseria m
39	210.5	10.5	595	1	TRPE_ARATH	P32068 arabidopsis
40	209.5	10.4	453	1	PABB_ECOLI	P05041 escherichia
41	209	10.4	462	1	TRPE_THETH	P05378 thermus aqu
42	208.5	10.4	491	1	TRPE_NETGO	Q9w400 neisseria g
43	208.5	10.4	492	1	TRPE_PSEAE	P20580 pseudomonas
44	201.5	10.0	621	1	TRPX_ARATH	P32069 arabidopsis
45	200.5	10.0	451	1	PABB_KLEAE	P12679 klebsiella

## ALIGNMENTS

```

RESULT 1
MENF_BACSU          STANDARD;          PRT:  471 AA.
ID  MENF_BACSU
AC  P23973; P23972;
DT  01-MAR-1992 (Rel. 21, Created)
DT  01-FEB-1996 (Rel. 33, Last sequence update)
DE  15-DEC-1998 (Rel. 37, Last annotation update)
DE  MENAQUINONE-SPECIFIC ISOCORISMATE SYNTHASE (EC 5.4.99.6).
GN  MENF OR ICSM.
OS  Bacillus subtilis.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group;
OC  Bacillus/staphylococcus group; Bacillus.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-168 / RB1.
RX  MEDLINE; 96144257.
RA  Rowland B., Hill K., Miller P., Driscoll J.R., Taber H.W.;
RT  "Structural organization of a Bacillus subtilis operon encoding
RT  menaquinone biosynthetic enzymes.";
RL  Gene 167:105-109(1995).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE; 98048467.
RA  Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT  "Sequencing and functional annotation of the Bacillus subtilis genes
RT  in the 200 kb rnb-dnaB region.";
RL  Microbiology 143:3431-3441(1997).
RN  [3]
RP  PRELIMINARY SEQUENCE FROM N.A.
RC  STRAIN-168 / RB1;
RX  MEDLINE; 92332443.
RA  Driscoll J.R., Taber H.W.;
RT  "Sequence organization and regulation of the Bacillus subtilis menB
RT  operon.";
RL  J. Bacteriol. 174:5063-5071(1992).
RN  [4]
RP  SEQUENCE OF 1-32 FROM N.A.
RX  MEDLINE; 88227858.
RA  Miller P., Mueller J., Hill K., Taber H.W.;
RT  "Transcriptional regulation of a promoter in the men gene cluster of
RT  Bacillus subtilis.";
RL  J. Bacteriol. 170:2742-2748(1988).
CC  -1- CATALYTIC ACTIVITY: CHORISMATE - ISOCORISMATE.
CC  -1- PATHWAY: MENAQUINONE BIOSYNTHESIS.
CC  -1- SIMILARITY: STRONG, TO OTHER ISOCORISMATE SYNTHASES; WEAK, TO
CC  TRPE AND PABB.
CC  -1- CAUTION: USED TO INCLUDE WHAT WAS CALLED 'MENR'.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).

```

DR EMBL; M74538; AAC37013.1; -.  
 DR EMBL; AF008220; AAC00223.1; -.  
 DR EMBL; M74521; AAAS0396.1; ALT\_SEQ.  
 DR EMBL; M74521; AAAS0397.1; ALT\_SEQ.  
 DR EMBL; M21320; AA22594.1; -.  
 DR EMBL; Z99119; CAB15061.1; -.  
 DR PIR; S27508; S27508.  
 DR PIR; S27507; S27507.  
 DR SUBTILIST; BG10682; MENF.  
 DR INTERPRO; IPR000350; -.  
 DR PFAM; PF00425; chorismate\_bind.1.  
 DR Menaglutone biosynthesis; Isomerase.  
 KW CONFLICT 12 E -> K (IN REF. 4).  
 SQ SEQUENCE 471 AA; 52811 MW; 94C047C549FC4CF9 CRC64;

Query Match 24.7%; Score 496.5; DB 1; Length 471;  
 Best Local Similarity 30.2%; Pred. No. 1.9e-29;  
 Matches 127; Conservative 77; Mismatches 161; Indels 55; Gaps 10;

OY 19 YAALETTERGAYFATPADDTLRFGALITAKTAQ----- 53  
 DB 46 YGAKYITGNRFWSDESELTIVGLKEAVFOTNOKNSRYEVEQWERFKTAHIE 105  
 OY 54 -----ALOGAVFGAQSFEDEYPOSE---LMAGFWPEVYVNTIADK-----IT 95  
 DB 106 EEKLGHSAGVPLFGGFSFDPCEBRSQMDHFSSEGDFVPLMTLTATGPFITVNRWS 165  
 OY 96 FGSD---IVSDTTWLAQF-VP--KOPN--YVTTSHVDEVMIEETEMLIDTLAIDQTL 147  
 DB 166 GGEDEAVALEGLKAPAEWMPDFKQEDQAVTAAAEELDKDWLKAEIATSOIK-EMKY 224  
 OY 148 AKVVGROOTLQSDTLRLAOLIRALA--EOANTYHVYLKRHDELFIATPERLVAMSGQ 206  
 DB 225 DKVVALRELLTFDPGFIQIEPVKLTLLDDQTSYFAIEQEGKTYGASPERLIKDDGT 284  
 OY 207 IATAVAGTSRRGTGADGADIALGEALLASQKNRIEHOYVVASITTRLQDVTSLKVPAMP 266  
 DB 285 VMSCLAGSIKRGVNEEDDRIGLELLNDEKMLEHDIYVGMIHNAFVSSCSEVEKPDGP 344  
 OY 267 SLKKKOVOLHTPTITGDIAHLSTAIYDRLHPTPALGQPREALYIATHEKTRPL 326  
 DB 345 VLYKTSVOHLFTPIVGOIRESASIFDLIEKLPALGSPQEKAVDVIETIEPMRGM 404  
 OY 327 FAGPIGYFADSGEYVVGIRSMVYNOTORATLFAAGIIVDSDAQOEFEETGLKEPM 386  
 DB 405 YAAPIGWDSQNGEFNAIRSGLIEGTAR--LFAGCIVDESEPISEYETQIKLKM 462

## RESULT 2

PCNA\_PSEAE STANDARD: PRT; 476 AA.  
 ID Q51508;  
 AC 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE SALICYLATE BIOSYNTHESIS ISOCHORISMATE SYNTHASE (EC 5.4.99.6).  
 GN PCNA.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAOI;  
 RX MEDLINE; 96086939.  
 RA Setino L., Reimann C., Baur H., Beyeler M., Visca P., Haas D.;  
 RT "Structural genes for salicylate biosynthesis from chorismate in  
 RT Pseudomonas aeruginosa.";  
 RL Mol. Gen. Genet. 249:217-228(1995).  
 CC -I- FUNCTION: INVOLVED IN THE CONVERSION OF CHORISMATE TO SALICYLATE  
 CC (PROBABLE).  
 CC -I- CATALYTIC ACTIVITY: CHORISMATE - ISOCHORISMATE.  
 CC -I- PATHWAY: SALICYLATE BIOSYNTHESIS.

CC -I- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO  
 CC TYPE AND PABB.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X82644; CAAS7969.1; -.  
 DR INTERPRO; IPR000350; -.  
 DR PFAM; PF00425; chorismate\_bind.1.  
 KW Isomerase.  
 SQ SEQUENCE 476 AA; 52071 MW; D0F6AFD9C9CFC9F5 CRC64;

Query Match 22.5%; Score 453; DB 1; Length 476;  
 Best Local Similarity 42.4%; Pred. No. 3.1e-26;  
 Matches 103; Conservative 38; Mismatches 98; Indels 4; Gaps 3;

OY 149 KYVFGQOQLQSDTLRLAOLIRAL-AEOANTYHVYLKRHDELFIATPERLVAMSGQI 207  
 DB 221 KYVALTQARPLGD-IEPQVIEHLRLQHADQLFACRGRNACFLGASPERLIRIRGEA 279  
 OY 208 ATAAVAGTSRRGTGADGADIALGEALLASQKNRIEHOYVVASITTRLQDVTSLKVPAMP 267  
 DB 280 LTHALGTLTARGDQADQEDARLGOALDSAKRHEHLYVEALITALEPSEVLEITDAG 339  
 OY 268 LKKKOVOLHTPTITGDIAHLSTAIYDRLHPTPALGQPREALYIATHEKTRGLE 327  
 DB 340 LKRLAVOHLPNTRIRALDAGILRLQALHPTPAVGPRSAALDYIRQHGMDRGW 399  
 OY 328 AGPIGFTADNNGSEFYVGRSMVYNOTORATLFAAGIIVDSDAQOEFEETGLKEPM 387  
 DB 400 AAPLGWLDGNGDFVLAIRSLD--TPGRGYLFAAGCGVGBSEPAHEVRETCLKLSMR 457  
 OY 388 QLL 390  
 DB 458 EAL 460

## RESULT 3

ENTC\_ECOLI STANDARD: PRT; 391 AA.  
 ID P10377;  
 AC 01-MAR-1989 (Rel. 10, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE ISOCHORISMATE SYNTHASE ENTC (EC 5.4.99.6).  
 GN ENTC.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE; 90236256.  
 RA Elkins M.F., Earhart C.F.;  
 RT "Opacity factor from group A streptococci is an apoproteolase.";  
 RL FEMS Microbiol. Lett. 56:35-40(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE; 89123153.  
 RA Ozenberger B.A., Brickman T.J., McIntosh M.A.;  
 RT "Nucleotide sequence of Escherichia coli isochorismate synthetase  
 RT gene entc and evolutionary relationship of isochorismate synthetase  
 RT and other chorismate-utilizing enzymes.";  
 RL J. Bacteriol. 171:775-783(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.



RC STRAIN-K12 / MG1655;  
 RX MEDLINE: 97426617.  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RT Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE OF 1-33 FROM N.A.  
 RX MEDLINE: 90230305.  
 RA Brickman T.J., Ozenberger B.A., McIntosh M.A.;  
 RT "Regulation of divergent transcription from the iron-responsive fepB-  
 RT entC promoter-operator regions in Escherichia coli.";  
 RL J. Mol. Biol. 212:669-682(1990).  
 RN [5]  
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.  
 RX MEDLINE: 90241936.  
 RA Liu J., Quinn N., Berchtold G.A., Walsh C.T.;  
 RT "Overexpression, purification, and characterization of isochorismate  
 RT synthase (EntC), the first enzyme involved in the biosynthesis of  
 RT enterobactin from chorismate.";  
 RL Biochemistry 29:1417-1425(1990).  
 CC -1- CATALYTIC ACTIVITY: CHORISMATE = ISOCHORISMATE.  
 CC -1- PATHWAY: ENTEROBACTIN BIOSYNTHESIS. ENTEROBACTIN IS AN IRON-  
 CC CHELATING COMPOUND INVOLVED IN TRANSPORTING IRON FROM THE  
 CC BACTERIAL ENVIRONMENT INTO THE CELL CYTOPLASM.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- INDUCTION: EXPRESSED UNDER CONDITIONS OF IRON STARVATION.  
 CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO  
 CC TRPE AND PABB.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M24142; AAA16100.1; -  
 DR EMBL: M36700; AAA18491.1; -  
 DR EMBL: AE000165; AAC73694.1; -  
 DR EMBL: X53274; CAA37371.1; -  
 DR PIR: J70497; SEYCIK.  
 DR PIR: S04323; S04323.  
 DR ECGENE: EG10261; ENTC.  
 DR INTERPRO: IPR000350; -  
 DR PFAM: PF00425; chorismate\_bind; 1.  
 DR Enterobactin biosynthesis; Iron transport; Isomerase.  
 KW CONFLICT 305 306 SG -> TA (IN REF. 1).  
 SQ SEQUENCE 391 AA; 42931 MW; 62862569DFC41AC4 CRC64;

Query Match 17.1%; Score 344; DB 1; Length 391;  
 Best Local Similarity 28.6%; Pred. No. 2.6e-18;  
 Matches 113; Conservative 65; Mismatches 153; Indels 64; Gaps 12;

QY 19 YAALETTEGAYPATPA-----DGLRFGIGAIATATQAALQGAVF-GAOSFDEQEXP 71  
 Db 27 YRSFTTSGCFARFDEPAVNDSPDFQKTLALFADAKAQGIRKPNVGAIPDPDQ-P 85  
 QY 72 QSLMAGFWFVPEVMTIADKLTTFGSDTVSDFTTWLAQGVPRQPNVTVTSHTYDEVDWT 131  
 Db 86 SS-----LTIPE-----SW--QSFSSROEKASARRFR----- 111  
 QY 132 ERTENTLIDTLAI-DGLTAKVVFGRQOTLQLS-----DTLRLOIIRALAEQNTYHVLKR- 186  
 Db 112 SGLNVERKAIPQOTTEEDMAAALITATPOVDKAVLSRLIDITDAIDSGLVERL 171  
 QY 187 -----HDEL-----FISATPERLVAMSGQIATAVAAGTSRRGTGADADIALGEA 231  
 Db 172 IAGNPVSYNPHVPLADGCVLLGASPELLLRKDGFRSSIFLAGSARRQPEVLDREAGNR 231

QY 232 LLASOKRIEHOYVVASITTRLODVTSLSKVPAMSLKKNQVOHLYPTIGDIAHLNV 291  
 Db 232 LLASOKRIEHELTQVMKEVLRERSSSELHPSSPOLITPTLWHLATPEGKANSQENA 291  
 QY 292 TATYDRDHPPTALGAGVPREALYIATHEKTPRGLFAGPIGTFYADNSGEFVGRMYV 351  
 Db 292 LTLACLHPTPALSGFPHQAATQVIALEPPDRFLPFGIYGVWCDSDENGENVYIRCAKL 351  
 QY 352 NOTORATLFGAGIVADSDAQOEYETGLKEFEP 386  
 Db 352 RENQVR--LFGAGIVPASSPLGEMRETVGLSTM 384  
 RESULT 4  
 ID DHBC\_BACSU STANDARD; PRT; 398 AA.  
 AC P45744;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE ISOCHORISMATE SYNTHASE DHBC (EC 5.4.99.6).  
 GN DHBC.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-168 / MARBURG;  
 RX MEDLINE: 96146538.  
 RA Rowland B.M., Taber H.W.;  
 RT "Duplicate isochorismate synthase genes of Bacillus subtilis:  
 RT regulation and involvement in the biosyntheses of menaquinone and  
 RT 2,3-dihydroxybenzoate.";  
 RL J. Bacteriol. 178:854-861(1996).  
 CC -1- CATALYTIC ACTIVITY: CHORISMATE = ISOCHORISMATE.  
 CC -1- PATHWAY: 2,3-DIHYDROXYBENZOATE BIOSYNTHESIS.  
 CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO  
 CC TRPE AND PABB.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U26444; AAC44631.1; -  
 DR EMBL: Z99120; CAB15189.1; -  
 DR SUBTILIST; BG11242; DHBC.  
 DR INTERPRO: IPR000350; -  
 DR PFAM: PF00425; chorismate\_bind; 1.  
 KW Isomerase.  
 SQ SEQUENCE 398 AA; 43446 MW; 3311F3B517C79BC CRC64;

Query Match 16.4%; Score 328.5; DB 1; Length 398;  
 Best Local Similarity 28.5%; Pred. No. 3.7e-17;  
 Matches 111; Conservative 64; Mismatches 163; Indels 51; Gaps 12;

QY 28 GAYF-ATPADDTLRFGIGATA-----TAKRQAALQGAVFAGQSFDEDEYQSE 74  
 Db 22 GAFFLASPHRYLLAKGICEIVPEADGQNMETLSGRIAEALROAKOSGQS----- 71  
 QY 75 LMAGFVPEVMTIADKLTTFGSDTVSDFTTWLA--QF--VPRQPNVTVTSH--VTDE 127  
 Db 72 -----RPLVGVAVPPDQYKAARLVVPEEVRWSGPIQGFHEEEOQAGHTYHKPVP 127  
 QY 128 VWIERTENTLIDTLAI-DGLTAKVVFGRQOTLQLS-DTLRLOIIRALAEQAN--TYHVL 184  
 Db 125 EDYKNGVEQGLARIA-DGLTSLKIVLSRLHLSPEPIQTDLRLHLOHNSHGTYFADY 183



RL J. Bacteriol. 171:4349-4354(1989).  
RN [8]  
RP IDENTIFICATION  
RX MEDLINE: 95075659.  
RA Borodovsky M., Rudd K.E., Koonin E.V.;  
RT "Intrinsic and extrinsic approaches for detecting genes in a  
bacterial genome.";  
RL Nucleic Acids Res. 22:4756-4767(1994).  
CC -1- CATALYTIC ACTIVITY: CHORISMATE - ISOCHORISMATE.  
CC -1- PATHWAY: MENADIOLONE BIOSYNTHESIS.  
CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO  
CC TRPE AND PAB.  
CC -1- CAUTION: REF.7 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO  
CC FRAMESHIFTS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z50849; CAA90702.1; -  
DR EMBL: U54790; AAC44303.1; -  
DR EMBL: AE000316; AAC75325.1; ALT\_INIT.  
DR EMBL: D90857; CAB22015.1; -  
DR EMBL: D90856; CAB22025.1; -  
DR EMBL: U58768; AAB02729.1; -  
DR EMBL: M21787; -; NOT\_ANNOTATED\_CDS.  
DR ECOGENE: EG12362; MENF.  
DR INTERPRO: IPR000350; -  
DR PIRAM: PF00425; chorismate\_bind; 1.  
DR Menadiolone biosynthesis; isomerase.  
KW CONFLICT 310 312 VLR -> YC (IN REF. 3).  
SQ SEQUENCE 431 AA; 48764 MW; D081724F3329FE47 CRC64;  
  
Query Match 15.8%; Score 317.5; DB 1; Length 431;  
Best Local Similarity 30.3%; Pred. No. 2.7e-16;  
Matches 106; Conservative 51; Mismatches 162; Indels 31; Gaps 10;  
  
QY 59 VFGAQSFEDEQEYQSELMAGFWFVEYVNTIADKIFGSDYSDPT-----TWLA 109  
DB 94 IMGLNMFDPQ-----GNLLPRLKMRGCGKATRLTLTSESSLDHDAIDAKFEITA 145  
QY 110 QFVPRKP-----NTVYTSHTVDEVMIERENTLIDTLAIDOTLAKVVFGRQOTLQSDT 163  
DB 146 TLVSIRKPLPGLHLTTREQHWPDKTGWTLQLELATKTLIA-EGELDKVLLARATDLHFASP 204  
QY 164 LRLAQITRALAE-QANTYHYVVKRHNDE-LFISATPERLVAMSGQIATAVAGTSRSGTD 221  
DB 205 VNAAMAAARRLNLCNYHYMAFDGEMAFGLSSPERLMMRRDKALRTEALAGTVANNPD 264  
QY 222 GADDIALGEALLSOKKRIRIEHOYVVASITRLODVTSLSKVPAMPSSLLKNKOYOHLTPI 281  
DB 265 DKQAQOLGEMLMDKKNOREMLVDEDICQLADTOTLTV-LPPQVLRLLKVKOHLRRCI 323  
QY 282 TGDIAHLSTVAIVRLHPTPALGVPREALAYIATHEKTRPGLGPIGFTADNSGE 341  
DB 324 WTSLSNKADVDYCL-HQLOPFAAVALGRLDLAGRFIARHEPFRKWMYAGSYLSLOS-E 381  
QY 342 FVVGISNMYNQTORRATLFAAGATVADSDAQOEYEETGKFERMOLK 391  
DB 382 FCVLSRSKISGNVVR--LYAGAGIVGSDPEQDEMOEIDNKAGLRILQL 429  
  
RESULT 7  
VIBC\_VIBCH STANDARD; PRT; 395 AA.  
ID VIBC\_VIBCH  
AC 007898;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE VIBRIOBACTIN-SPECIFIC ISOCHORISMATE SYNTHASE (EC 5.4.99.6).  
GN VIBC.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN: EL TOR / L0015;  
RX MEDLINE: 98037504.  
RA Wyckoff E.E., Stoeckner J.A., Reed K.E., Payne S.M.;  
RT "Cloning of a Vibrio cholerae vibriobactin gene cluster:  
RT identification of genes required for early steps in siderophore  
biosynthesis.";  
RL J. Bacteriol. 179:7055-7062(1997).  
CC -1- CATALYTIC ACTIVITY: CHORISMATE - ISOCHORISMATE.  
CC -1- PATHWAY: VIBRIOBACTIN BIOSYNTHESIS. VIBRIOBACTIN IS AN IRON-  
CC CHELATING COMPOUND INVOLVED IN TRANSPORTING IRON FROM THE  
CC BACTERIAL ENVIRONMENT INTO THE CELL CYTOPLASM.  
CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO  
CC TRPE AND PAB.  
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U52150; AAC45925.1; -  
DR INTERPRO: IPR000350; -  
DR PIRAM: PF00425; chorismate\_bind; 1.  
DR Iron transport; isomerase.  
KW SEQUENCE 395 AA; 43594 MW; 2FFAF87DD948C78B CRC64;  
  
Query Match 15.4%; Score 308.5; DB 1; Length 395;  
Best Local Similarity 28.5%; Pred. No. 1.1e-15;  
Matches 109; Conservative 57; Mismatches 158; Indels 59; Gaps 11;  
  
QY 30 YEATPADDTLRFGIG-----AIATAKTA---QALOGAVFGAQSFEDEYQSELMAGFW 80  
DB 28 FFAASNSMLGCGVAHAFQQAIPFAELANAKQLLOQA-----KRDECNDP---LFGI- 78  
QY 81 FVPEVNTIADKITFGSDYSDPTWLAQFVPKPQPTVTSHTVDEVMIERENTLIDT 140  
DB 79 -----VFPKPPTPRMIPRLTVYSSPRLNRPRAHLRLQV-----AKLSS 119  
QY 141 LAIDQ-----TLAVVFGRQOTLQSDTLRLAQILR---ALAEQANTYHV 182  
DB 120 PSGEOYKQGVSHLNMENHSGLSKVLRSRALEITEDEIALPTLLRSILAIHNGYFAA 179  
QY 183 VLKRHDELFISATPERLVAMSGQIATAVAGTSRSGTDGADDIALGELLASOKNRIE 242  
DB 180 SLDEQRKL-IGASPELLVAKRGNLTLSNPLAGSRPSODAOENORRASSLNTAKDHEH 238  
QY 243 QYVVASITRLODVTSLSKVPAMPSSLLKNKOYOHLYPTIGDIA-AHLSTVAIVRLHPT 301  
DB 239 GLVVEEYERIMSRCNMLYTPMVPVSYTEETMLHLSTLBEQVSDPEFCALQVAAADLHPT 298  
QY 302 PALGCVPREALAYIATHEKTRPGLGPIGFTADNSGEFVVGIRSNYVYNTORRATLF 361  
DB 299 PAVGCFPRESAYQAIREFLEEDRDYFTGMVGWCARGNGEYVYIRCAEVSQMK--LF 356  
QY 362 AGAGIVADSDAQOEYEETGKFE 384  
DB 357 AGAGIVDESLPQSELETGAKMK 379  
  
RESULT 8  
TRPE\_SULSO  
ID TRPE\_SULSO STANDARD; PRT; 421 AA.

```

AC Q06128;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27).
GN TRPE.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 49255 / MT4;
RA MEDLINE: 93106970.
RX Tullino M.L., Scarano G., Marino G., Sanna G., Cubellis M.V.;
RT "Tryptophan biosynthesis genes trpEGC in the thermoacidophilic
  archaeobacterium Sulfolobus solfataricus.";
RL J. Bacteriol. 175:299-302(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 49255 / MT4;
RA Tullino M.L., Cubellis M., Sanna G., Marino G.;
RT "The tryptophan operon in Sulfolobus solfataricus.";
RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE: 99380543.
RA Knoechel T., Ivens A., Hester G., Gonzalez A., Bauerle R.,
  Wilmanns M., Kirschner K., Jansonsius J.N.;
RT "The crystal structure of anthranilate synthase from Sulfolobus
  solfataricus: functional implications.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9479-9484(1999).
CC -1- CATALYTIC ACTIVITY: CHORISMATE + L-GLUTAMINE = ANTHRANILATE +
  PYRUVATE + L-GLUTAMATE.
CC -1- PATHWAY: FIRST STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC -1- SUBUNIT: TETRAMER OF TWO COMPONENTS I AND TWO COMPONENTS II.
CC -1- MISCELLANEOUS: COMPONENT I CATALYZES THE FORMATION OF ANTHRANILATE
  USING AMMONIA RATHER THAN GLUTAMINE, WHEREAS COMPONENT II PROVIDES
  GLUTAMINE AMIDOTRANSFERASE ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE ANTHRANILATE SYNTHASE COMPONENT I
  FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M98048; AAA73379.1; -
DR SRSB: Z50014; CAA90311.1; -
DR PIR: A40635; A40635.
DR PDB: 1ODL; 18-AUG-99.
DR INTERPRO: IPR000350; -
DR PFM: PF00425; chorismate_bind.1.
DR PRINTS: PR00095; ANTSNTHASEI.
DR Tryptophan biosynthesis; Lyase; 3D-structure.
KW SEQUENCE 421 AA; 47736 MW; 0E5B9E776E298C61 CRC64;
SQ

```

```

DB 169 EYIR-----SGYIFQVVLRSFRYIFSGDPLRIYNNLRRI--NPSPYMFLKFD 215
QY 188 DELFISATPERLIAMGGGQATATAVAGTSRGRGDADIALGALLASCKNRREHQYVA 247
DB 216 EKYLGSSPELLRVDONIVETYPDIAGTRPGADQEDDLLELMNSEKDKAHEMLVD 275
QY 248 SITTRQDVVT--SLKVPAMPSILKKQVOHLTPPTGDIAMHLSYATAYDRHPTPALG 305
DB 276 LANLDGKVCVPGCTVAPPELMVYEKSHOHIYSYIGTLKKRYNNALNVLSATFPAGTYS 335
QY 306 GVPREALYIATHEKTPNGLEFAGPIGYFTADNSGEFVGIRSMYVNGTORATLPAGAG 365
DB 336 GRKPPAMANNIEETLEEKYKGPYAGVGFISADGNAEFALIRPAFLINKELLR--IHAGAG 393
QY 366 IVADSDAQOEYEETGLKFEPMR 387
DB 394 IYDSNPSESEFEETHEKALK 415

RESULT 9
TRPE:HALVO STANDARD; PRT; 523 AA.
ID TRPE:HALVO
AC P33975;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27).
GN TRPE.
OS Halobacterium volcanii (Haloferax volcanii).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloferax.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WED11;
RX MEDLINE: 92165748.
RA Lam W.L., Logan S.W., Doolittle W.F.;
RT "Genes for tryptophan biosynthesis in the halophilic archaeobacterium
  Haloferax volcanii: the trpFEG cluster.";
RL J. Bacteriol. 174:1694-1697(1992).
CC -1- CATALYTIC ACTIVITY: CHORISMATE + L-GLUTAMINE = ANTHRANILATE +
  PYRUVATE + L-GLUTAMATE.
CC -1- PATHWAY: FIRST STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC -1- SUBUNIT: TETRAMER OF TWO COMPONENTS I AND TWO COMPONENTS II (BY
  SIMILARITY).
CC -1- MISCELLANEOUS: COMPONENT I CATALYZES THE FORMATION OF ANTHRANILATE
  USING AMMONIA RATHER THAN GLUTAMINE, WHEREAS COMPONENT II PROVIDES
  GLUTAMINE AMIDOTRANSFERASE ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE ANTHRANILATE SYNTHASE COMPONENT I
  FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M83788; AAA73177.1; -
DR PIR: A42301; A42301.
DR INTERPRO: IPR000350; -
DR PFM: PF00425; chorismate_bind.1.
DR PRINTS: PR00095; ANTSNTHASEI.
DR Tryptophan biosynthesis; Lyase.
KW SEQUENCE 523 AA; 55920 MW; 976566797636A6C CRC64;
SQ

```

```

Query Match 14.4%; Score 289; DB 1; Length 421;
Best Local Similarity 24.9%; Pred. No. 3.3e-14;
Matches 95; Conservative 72; Mismatches 163; Indels 52; Gaps 11;

```

```

Query Match 13.9%; Score 278.5; DB 1; Length 523;
Best Local Similarity 28.9%; Pred. No. 2.6e-13;
Matches 79; Conservative 44; Mismatches 115; Indels 35; Gaps 5;

```

OY	174	AEOANTYVHVLRKHDELPTSAPEELVAMSGQIITAAVAGRSRGTCGADIALGEALL	233
OY	293	--NSPFIWFLRHGDRRVVGASPELTVSRGDRVVNFPJAGTCKRGSCEVEDRRRLAGELL	350
OY	234	ASOKNRIRHGVVMSITTRLADVTT--SLKVPAMPILKNKOVHLPIPTIGDIAHLSTV	291
Db	351	ADAKRAHHTMLVLDGRNDVRVSTRPSVRREDFMSIIKYSHVOIHTESTYSTDLADADA	410
OY	292	TALVDRLHPPLAGVPREALYYATHEKTTPRGLFAPCIPTGYFTADNGSEFYVGRSMYV	351
Db	411	FDATRPATFPACTLTGCAPRVAMEIIDLEAEPRGVYGCVGGVYSWTGDAVAIVRTATV	470
OY	352	NQ--TGRRATLEFACGIYADSDAOEYETGTGK	382
Db	471	DSCGADAIIVRAGIVADSPDTAVEYEETEOK	503
RESULT	10		
MENF_HAEIN	ID	MENF_HAEIN	STANDARD; PRT; 430 AA.
AC	P44613;		
DT	01-NOV-1995	(Rel. 32, Created)	
DT	01-NOV-1995	(Rel. 32, Last sequence update)	
DT	01-NOV-1995	(Rel. 32, Last annotation update)	
DE	MENAQUINONE-SPECIFIC ISOCHORISMATE SYNTHASE (EC 5.4.99.6).		
GN	MENF OR H10285.		
OS	Haemophilus influenzae.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;		
CC	Haemophilus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-RD / KW20.		
RX	MEDLINE: 95350630.		
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,		
RA	Ketlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,		
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,		
RA	Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,		
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,		
RA	Uettermann T.R., Hanna M.C., Nguyen D.T., Sandek D.M., Brandon R.C.,		
RA	Fine L.D., Fitchman J.L., Fuhrmann J.L., Georgagen N.S.M.,		
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,		
RA	Venter J.C.;		
RT	"Whole-genome random sequencing and assembly of Haemophilus		
RT	influenzae Rd."		
RL	Science 269:496-512(1995).		
CC	-1- CATALYTIC ACTIVITY: CHORISMATE = ISOCHORISMATE.		
CC	-1- PATHWAY: MENAQUINONE BIOSYNTHESIS.		
CC	-1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO		
CC	TYPE AND PAAB.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on ways		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a> )		
CC	or send an email to <a href="mailto:license@isb-sdb.ch">license@isb-sdb.ch</a> .		
CC	-----		
DR	EMBL: U32714; AAC21947.1; -		
DR	TIGR: H10285; -		
DR	INTERPRO: IPR000350; -		
DR	PFAM: PF00425; chorismate_bind.1.		
DR	MegaQuinone Biosynthesis; Isomerase.		
SQ	SEQUENCE 430 AA; 48466 MW; 3BA11FB862A1B38 CRC64;		

Query Match	13.8%	Score 278;	DB 1;	Length 430;
Best Local Similarity	27.8%;	Pred. No. 2.2e-13;		
Matches 101; Conservative	66;	Mismatches 144;	Indels 52;	Gaps 17,
QY	52	ACALGAVFGAQSFD	---OEYPOSE---	LMAGFWF-----VPEVAV-----TINA 91

```

Db 68 ALAALGAN--QSPSOLNLAQEFIEESGFLVGLQJQGNAGQFLPKMLVEDQNKGL-- 122
OY 92 DKTF-----GSDYVSDFTW---LAEVFKQPMVTTSHTVDEY--DMIERENLI 138
Db 123 --VSFEVNEOSADDTLALHTLFENLTALTSALPQOILHTLELRANERTMCDWNOA--LV 178
OY 139 DTLAIDQFLAVVGGROQTQDSTLRLOQIRALAOAN--TTHVY-LKRHDELFTSAT 195
Db 179 EIKSE--LTKIVLANETTFHLKQAINAYDFL-RESERKQGGCHFLMAENSSVFEVGT 235
OY 196 PERLVAMSGOIAATAAAGVSRROTDGADIDALGEALLASQKNRIEHQVYVAISITRLOD 255
Db 236 PERLFARERYNLLTEALAGTASVSESEEEYQSQNMWLLNDEKNLKEMWLVVEDISQNLRK 295
OY 256 VTTSLKVPAMPSSLKNVOHLYTPITGDIANAHLSTVAIVYDRCLHPALGCVPREALVY 315
Db 296 QVESFEDYSNV-ELKPLRKVYOHLLKIRIRANLTATHADVADNIIKAIHPTAAVSGLPQOQAKMI 354
OY 316 IATHKTRGLFAPRIGITFTADNSGFEVYGRISMYVMOTQRRLATFLFGAGAIVADSDAQOE 375
Db 355 LSEIETFDRGVAGTLGVN-SDVCSEFCVAILRSFIE--GHRIRVFGAGIIVAGSQPLEE 411
OY 376 YEE 378
Db 412 WKE 414

```

TRPE_LACIA	11	RESULT
ID	TRPE_LACIA	STANDARD: PRT: 456 AA.
AC	002001;	
DT	01-APR-1993 (Rel. 25, Created)	
DT	01-APR-1993 (Rel. 25, Last sequence update)	
DT	30-MAY-2000 (Rel. 39, Last annotation update)	
DE	ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27).	
NC	TRPE.	
OS	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).	
OC	Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	
OC	Lactococcus.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-11403;	
RX	MEDLINE: 93015708.	
RA	Bardowski J., Ehrlich S.D., Chopin A.;	
RT	"Tryptophan biosynthesis genes in Lactococcus lactis subsp. lactis.";	
RL	J. Bacteriol. 174:6563-6570(1992).	
CC	-I CATALYTIC ACTIVITY: CHORISMATE + L-GLUTAMINE = ANTHRANILATE +	
CC	PYRUVATE + L-GLUTAMATE.	
CC	-I PATHWAY: FIRST STEP IN BIOSYNTHESIS OF TRYPTOPHAN.	
CC	-I SUBUNIT: TETRAMER OF TWO COMPONENTS I AND TWO COMPONENTS II (BY	
CC	SIMILARITY).	
CC	-I MISCELLANEOUS: COMPONENT I CATALYZES THE FORMATION OF ANTHRANILATE	
CC	USING AMMONIA RATHER THAN GLUTAMINE, WHEREAS COMPONENT II PROVIDES	
CC	GLUTAMINE AMIDOTRANSFERASE ACTIVITY.	
CC	-I SIMILARITY: BELONGS TO THE ANTHRANILATE SYNTHASE COMPONENT I	
CC	FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL, M87483; AAA25223.1; -.	
DR	PIR, S35124; S35124.	
DR	INTERPRO: IPR000350; -.	
DR	PFAM: PF00425; chorismate_bind.1.	
DR	PRINTS: PRO0095; ANTSNTHASE1.	
DR	TRYPTOPHAN biosynthesis; Lyase.	
DR	SEQUENCE 456 AA; 51837 MW; 8CDECD949BD55C3 CRC64;	









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2000, 02:17:04 ; Search time 54.8 Seconds  
(without alignments)  
676.451 Million cell updates/sec

Title: US-08-774-104A-2  
Perfect score: 2009  
Sequence: 1 MMTYHETRALAQSDLOQLYA.....ETGLKFEPMROLLDYNIVE 397

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues  
Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL.14:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.oranelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp rodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1998.5	99.5	398	4	O9UIA3	O9uia3 homo sapien
2	383	19.1	474	2	P74053	P74053 synechocyst
3	360	17.9	580	10	O9ZPC0	O9zpc0 catcharanthu
4	354.5	17.6	372	2	O05851	O05851 mycobacteri
5	344	17.1	395	2	P77099	P77099 escherichia
6	307.5	15.3	503	10	O81522	O81522 arabidopsis
7	307.5	15.3	503	10	O9S7H8	O9s7h8 arabidopsis
8	304.5	15.2	391	2	P95475	P95475 pseudomonas
9	295.5	14.7	449	8	O9TM09	O9tm09 cyanidulum c
10	251.5	12.5	434	4	O9V1G5	O9v1g5 pyrococcus
11	227	11.3	502	2	O9S2U2	O9s2u2 streptomyce
12	219	10.9	719	2	P72539	P72539 streptomyce
13	218.5	10.9	489	3	O94582	O94582 schizosacch
14	217.5	10.8	434	2	O9X9I8	O9x9i8 yersinia en
15	216.5	10.8	434	2	O9Z396	O9z396 yersinia pe
16	216	10.8	434	2	O9Z6F1	O9z6f1 yersinia pe
17	215.5	10.7	638	2	O9RXV9	O9rxv9 deinococcus
18	215	10.7	503	2	O9ROG2	O9rog2 rhodobacter
19	214.5	10.7	473	2	O9RTH3	O9rth3 deinococcus

20	210.5	10.5	364	3	P78905	P78905 schizosacch
21	202.5	10.1	522	2	O9R044	O9r044 buchnera ap
22	202.5	10.1	613	10	O41156	O41156 ruta graveo
23	199.5	9.9	450	2	P71758	P71758 mycobacteri
24	199.5	9.9	608	10	O41155	O41155 ruta graveo
25	195	9.7	616	10	O81533	O81533 nicotiana t
26	189.5	9.4	606	10	O9XJ29	O9xj29 oryza sativ
27	187.5	9.3	669	2	O56151	O56151 streptomyce
28	187.5	9.3	979	5	O9Y1K0	O9y1k0 plasmodium
29	181.5	9.0	577	10	O9XJ30	O9xj30 oryza sativ
30	177.5	8.8	917	10	O9ZV26	O9zv26 arabidopsis
31	171	8.5	391	2	O9RMT7	O9rmt7 streptomyce
32	167	8.3	518	2	O9R5Z6	O9r5z6 corynebacte
33	161.5	8.0	718	3	O94277	O94277 schizosacch
34	153	7.6	567	2	O9ZME2	O9zme2 helicobacte
35	151	7.5	328	2	O57527	O57527 haemophilus
36	150	7.5	458	2	O05591	O05591 mycobacteri
37	149	7.4	559	2	O25066	O25066 helicobacte
38	132	6.6	348	2	O67217	O67217 aquifex aeo
39	131.5	6.5	376	2	O46681	O46681 escherichia
40	131.5	6.5	336	2	O46682	O46682 escherichia
41	131	6.5	336	2	O57139	O57139 escherichia
42	131	6.5	336	2	O57446	O57446 escherichia
43	129	6.4	336	2	O46683	O46683 escherichia
44	124	6.2	623	2	O30589	O30589 streptomyce
45	113	5.6	824	3	O9Y7F1	O9y7f1 aspergillus

#### ALIGNMENTS

RESULT 1  
ID O9UIA3 PRELIMINARY; PRT; 398 AA.  
AC O9UIA3;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE NEUTRAL SPHINGOMYELINASE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TTSUE-KIDNEY;  
RA Chatterjee S., Han H., Rollins S., Cleveland T.;  
RT "Molecular cloning, characterization, and expression of a novel human  
RT neutral sphingomyelinase.";  
RL J. Biol. Chem. 274:37407-37412(1999).  
DR EMBL: AF069740; AAF19052.1; -.  
DR INTERPRO: IPR000350; -.  
DR PFM: PFM0425; Chorismate-bind; 1.  
DR PRINTS: PRO0095; ANTSNTHASE1.  
SQ SEQUENCE 398 AA; 43590 MW; 0248D89974C7BC1 CRC64;

Query Match 99.5%; Score 1998.5; DB 4; Length 398;  
Best local Similarity 99.7%; Pred. No. 1.6e-146;  
Matches 397; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 MMTYHETRALAQSDLOQLYALETTERGAFARPADDTLRFGIGATATKTAQAQGA-V 59  
DB 1 MMTYHETRALAQSDLOQLYALETTERGAFARPADDTLRFGIGATATKTAQAQGA-V 60  
OY 60 FGAQSEDEEYPOSELNAGFWFVPEVWVTTAADKITFGSDTVSPFTWLAQFVKKPQNTV 119  
DB 61 FGAQSEDEEYPOSELNAGFWFVPEVWVTTAADKITFGSDTVSPFTWLAQFVKKPQNTV 120  
OY 120 TTSHTVDEVDVMIERTENLIDTLAIDQTLAKVVFGRQDTQLSDTLRLAQITRALAEQANT 179  
DB 121 TTSHTVDEVDVMIERTENLIDTLAIDQTLAKVVFGRQDTQLSDTLRLAQITRALAEQANT 180  
OY 180 YHVYVLRKHDELFSATPERLVAMSGGQIATNAAGTSRRGTDCADDIALGELLASQKNR 239

```
|||||
Db 181 YHVLKRDELFI SATPERLVAMSGQIATAVAGTSRRGTGADIALGEALLASOKNR 240
Qy 240 IEHOYVVASITTRLODVTTSLKVPAMPISLKKKOVQHLYTPTTGDIHAALSTAIYDRH 299
Db 241 IEHOYVVASITTRLODVTTSLKVPAMPISLKKKOVQHLYTPTTGDIHAALSTAIYDRH 300
Qy 300 PTPALGVPREALYYIATHEKTPRGLEFAGPIGYFTADNSGEFVVGIRSMYNOTORRAT 359
Db 301 PTPALGVPREALYYIATHEKTPRGLEFAGPIGYFTADNSGEFVVGIRSMYNOTORRAT 360
Qy 360 LFGAGIYVADSDAOQEYEETGLKFEPMROLKDYNHVE 397
Db 361 LFGAGIYVADSDAOQEYEETGLKFEPMROLKDYNHVE 398

RESULT 2
P74053 PRELIMINARY; PRT; 474 AA.
ID P74053
AC P74053;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE ISCHORI SMATE SYNTHASE.
GN ENTC.
OS Synechocystis sp. (Strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97061201.
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Sugiyura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nanno K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tanaka S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90911; BAA18129.1; -.
DR INTERPRO; IPR000350; -.
DR PFM; PF00425; choris mate_bind; 1.
DR PRODOM; PD000779; -.
SQ SEQUENCE 474 AA; 52667 MM; 18F8BDB9E95C492C CRC64;

Query Match 19.1%; Score 383; DB 2; Length 474;
Best Local Similarity 31.5%; Pred. No. 1.2e-21;
Matches 112; Conservative 66; Mismatches 150; Indels 28; Gaps 9;

Qy 59 VEGASPFDEOEY--POSELMAGFWFVPEVMYT-----IAADKTFPGSDTVSDFETW 107
Db 123 IFCGSFEDRPSNRNRPANSPFLFQIQVYKTSQHCLLSWQSLDGNVTWTLVDPIGIM 182
Qy 108 LAQFVYKOP--NTVTTSHTVD-----EVDMIERT-ENLIDTLAIDOTLAKVVGKROQT 157
Db 183 LSAIRAQPARDTHTPPSVVAKPRLTGEVAKLSKAIASLSEIA--QQRLSVVLATAID 241
Qy 158 LQLSTLTLAQITIRALAEQANTYHYVL--KRDELFI SATPERLVAMSGQIATAVAGT 215
Db 242 LDYGSRLVAVHGLQRLROOYGDICYLFSGWNGGDCFVGASPERLSLHNOOLVTDALAGS 301
Qy 216 SRRGTGADADIALGEALLASOKNRIEHOYVVASITTRLODVTTSLKVPAMPISLKNKOY 275
Db 302 APRDNDVVGDRLOGLLHNRKELREHQAVIDYLIQRLALAGLSQASSL-KLKLKLANIQ 360
Qy 276 HLXPTPTGDIHAALSVTAIVDLRHPRLALGVPREALYYIATHEKTPRGLEFAGPIGYFT 335
Db 361 HLMTQIQAPLRPHIRPLALVQOLHPTPAVAGVPAIAEDLIRRHETFPBNLTAAPLGMWD 420
Qy 336 ADNSEGFVVGIRSMYNOTORRATLFGAGIYVADSDAOQEYEETGLKFEPM-ROLL 390
Db 421 SEGNAEFTIVGIRSLAL--SRNRLAYAGAGIYVADSDPLKEVAEIELKLOTLMRSLL 474
```

```
RESULT 3
Q92PCO PRELIMINARY; PRT; 580 AA.
ID Q92PCO
AC Q92PCO;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE ISCHORI SMATE SYNTHASE (EC 5.4.99.6).
OS Catharanthus roseus (Rosa periwinkle) (Madagascar periwinkle).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Asteridae; eusterids I; Gentianales;
OC Apocynaceae; Catharanthus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-G. DON (MADAGASCAR PERIWINKLE);
RA van Tegelen L.J.P., Moreno P.R.H., Croes A.F., Verpoorte R.,
RA Willems G.J.;
RT "Purification and characterization of isochoris mate synthase from
RT elicited cell cultures of Catharanthus roseus (L.) G. Don.";
RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ006065; CAA06837.1; -.
DR INTERPRO; IPR000350; -.
DR PFM; PF00425; choris mate_bind; 1.
DR PRODOM; PD000779; -.
ISomerase.
SQ SEQUENCE 580 AA; 64072 MM; 83F9633B4067312D CRC64;

Query Match 17.9%; Score 360; DB 10; Length 580;
Best Local Similarity 29.9%; Pred. No. 9.4e-20;
Matches 108; Conservative 65; Mismatches 156; Indels 32; Gaps 11;

Qy 60 FGAQSFDEOEY--POSELMAGFWF-VPEY-----MTYIAD---KITFGS----- 98
Db 216 YCAIRDPARPHIAPKFAKGFVFPVQVEFDELHSSMIATVMDNALSTLYQOAIYA 275
Qy 99 -----DVSDFETWTLAQFVKOPNRYTTSYHVDVDMIERTELIDTLAIDOT-LAKVVF 152
Db 276 LQTTMEQVSSIVSKLRQDY-SHTSLVSKANIPDRPSWDLTLNRVLEIGNKISPLTKVL 334
Qy 153 GQO-QTLQSDPTLRALQIIRALAEQANTYHYVVLKRHD-ELFTSATPERLVAMSGQIATA 210
Db 335 ARRSQVITTSIDIDPLAWLSFRKADKDAVOFCLQPREAPAFIGNPDEQFGDQTLVFE 394
Qy 211 AVAGTSRRGTGADADIALGEALLASOKNRIEHOYVVASITTRLODVTTSLKVPAMPISL 270
Db 395 ALAATRRARESSDLDLQMAHDLFSSPKDNHERAIVRENTROKLDICTSVETEPKMSYRK 454
Qy 271 NKVOYHLVPIRTGDIHAALSVTAIVDLRHPRLALGVPREALYYIATHEKTPRGLEFAGP 330
Db 455 LKRIQHLIARFGRGLRSEDEKRLSSLHPTPAVCGFPHEDKRKFLAENEMDRGLIYAGP 514
Qy 331 IGYFTADNSGEFVVGIRSMYNOTORRATLFGAGIYVADSDAOQEYEETGLKFEPMROL 390
Db 515 VGFPGAGS-DESVGIRSLALIK-DAGALITYAGLVESDPALENGELKASQFMKLM 572
Qy 391 K 391
Db 573 K 573
```

```
RESULT 4
O05851 PRELIMINARY; PRT; 372 AA.
ID O05851
AC O05851;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE ENTC.
GN ENTC OR RV3215 OR MTCY07D11.11C.
OS Mycobacterium tuberculosis
```

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 MDLINE: 98295987.  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,  
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 Davies R., Devlin K., Felwell T., Genies S., Hamlin N., Holroyd S.,  
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 Rutter S., Seeger K., Skelton S., Squares S., Squires R., Stalton J.E.,  
 Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 DR EMBL: Z95120; CAB08301.1; -.  
 DR TUBERCULIST: RV3215; -.  
 DR INTERPRO: IPR000350; -.  
 DR PRAM: PF00425; Chorismate\_bind; 1.  
 DR PRODOM: PD000779; -; 1.  
 SQ SEQUENCE 372 AA; 38937 MW; E2C1480BC281DF3 CRC64;

Query Match 17.6%; Score 354.5; DB 2; Length 372;  
 Best Local Similarity 36.6%; Pred. No. 1.3e-19;  
 Matches 102; Conservative 35; Mismatches 129; Indels 13; Gaps 6;

QY 116 PNTVTSHTVDEVDWERTENLIDTL-AIDQTLAKVFGROQTQLSDTLRLAQITRALA 174  
 Db 85 PKYVVAALPPPADYLTRIGRDLAFAFGDLKVVLAARVQLTADARVLLRLTV 144  
 QY 175 EQANTYIVVL-----KHNDE---LFISATPERLYAMSGGQIATAVAGTSRRGTGADDI 226  
 Db 145 VADPTAYGVLDVLTSAQNDGTALVGASPELLVARSGNRYMCKPFGASAPRADPKIDA 204  
 QY 227 ALGALLASQKNRIEHQYVASITTRLODVTSLSKVPAMPSLKNKOYOHLYPTIGDIA 286  
 Db 205 ANAALLASSKKNRHHQVLDVTRVVALEPCEDLTTPRQPLNTAAVWHCTAITGRL- 263  
 QY 287 AHSVTAI--VDRLHPTPALGVPREALYYIATHEKTPRGLFAPGIGYFADNSGEFVV 344  
 Db 264 RNISTTAIDLALHPTPAAGVPTKATIELALE-GDRFGYGAAGVCMCGRDRGHVYV 322  
 QY 345 GINSMYVNOTQRATLTFAGAGIADSDAQOEYEETGLKF 383  
 Db 323 SIRCAQLSADRRAALAHAGGIVAESDPDELEETTKF 361

RESULT 5  
 P77099 PRELIMINARY; PRT; 395 AA.  
 AC P77099;  
 DT 01-FEB-1997 (TREMblrel. 02, Created)  
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)  
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
 DE ISOCORISMATE SYNTHASE.  
 GN ENTC.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,  
 Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,  
 Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U82598; AAB40793.1; -.  
 DR INTERPRO: IPR000350; -.  
 DR PRAM: PF00425; Chorismate\_bind; 1.  
 DR PRODOM: PD000779; -; 1.  
 SQ SEQUENCE 395 AA; 43422 MW; 47D925F1953DB933 CRC64;

Query Match 17.1%; Score 344; DB 2; Length 395;  
 Best Local Similarity 28.6%; Pred. No. 9.2e-19;  
 Matches 113; Conservative 65; Mismatches 153; Indels 64; Gaps 12;

QY 19 YAALETTFEGAYFAPRA-----DITLFGICATATTAQALQAVF-GAQSFEQERYP 71  
 Db 31 YRSFTTSCGFARFDEPAGSDSPSPFOOKLAALPADAKAGIKNPVWGAIPDPQ-P 89  
 QY 72 QSELMAGFWPEVENVYIAADKIFGSDTVSDFTWLQOFVQKONVTYTSHTVDEVMI 131  
 Db 90 SS-----LYIPE-----SW-QSPSRQKQASARFTR----- 115  
 QY 132 ERTENLIDTLAI-DQTLAKVFGROQTQLS---DITLRLAQITRALAQANTYHVILK- 186  
 Db 116 QSGLNVVERQAIPEQTTEQWVARAALTAAPQVKVILSLRIDITTPDAIDSGVLLRL 175  
 QY 187 -----HDEL-----FISATPERLYAMSGGQIATAVAGTSRRGTGADDIALGEA 231  
 Db 176 IAQNEVSYNFPVPLADGVLGASPELLLRKDERFSSIPLAGSARQPDVLDREAGNR 235  
 QY 232 ILASQKNRIEHQYVASITTRLODVTSLSKVPAMPSLKNKOYOHLYPTIGDIAHLSV 291  
 Db 236 ILASEKDHHEHLYVQAKEVLRERSSELHVPSSQTLITPILHILATPPEKANSQENA 295  
 QY 292 TAIYDLRHLPTPALGVPREALYYIATHEKTPRGLFAPGIGYFADNSGEFVVYIRMYV 351  
 Db 296 LTLACLHPTPALSGFPQAATQVIALEPDRLELFGIVGCDSENGENVVYTRCAKL 355  
 QY 352 NOTQRATLTFAGAGIADSDAQOEYEETGLKFEP 386  
 Db 356 RENQVR--LFAGAGIVPASSPLGIEWETGVKLSTM 388

RESULT 6  
 O81522 PRELIMINARY; PRT; 503 AA.  
 AC O81522;  
 DT 01-NOV-1998 (TREMblrel. 08, Created)  
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
 DE ISOCORISMATE SYNTHASE.  
 OS Arabidopsis thaliana (mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;  
 OC Brassicaceae; Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Meng H., Peter G., Pullman G.;  
 RT "Arabidopsis isochorismate synthase (AB-ics1) cDNA sequence."  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF078080; AAC97926.1; -.  
 DR INTERPRO: IPR000350; -.  
 DR PRAM: PF00425; Chorismate\_bind; 1.  
 DR PRODOM: PD000779; -; 1.  
 SQ SEQUENCE 503 AA; 55359 MW; 50FAE4A2DE862742 CRC64;

Query Match 15.3%; Score 307.5; DB 10; Length 503;  
 Best Local Similarity 28.2%; Pred. No. 8.6e-16;  
 Matches 106; Conservative 59; Mismatches 158; Indels 53; Gaps 11;

QY 21 ALETTTEFGA-YFATPADDTLRFGIGAITAKTA-----QALQGVFEGAGSFEQERYPQS 73  
 Db 147 AVENEHPGAEFVSFQVEFNFGSSMLAATIAMDELSTWLENALIELQ----- 196  
 QY 74 ELAMGFVPEVWYTIADKTTTPGSDTVSFTTWLAQFVPKQPNVTYTSHTVDEVMI 133  
 Db 197 -----ETMLQVSSVVKLNRNSIG-----VSYSKNHVPTRGAYRPA 233  
 QY 134 TENLIDTL-AIDQTLAKVFGROQTQLSDT-----LRLAQITRALAQANTYHVILKRN 187

```
Db 234 VEKALEMINOKSSPLNKVVLARNRI-TTDTIDPIAMLAQLOR---EGHDVAFCLOPP 289
OY 188 D-ELFISATPERLVAMSGGQIATAVAGTSRRGTGADGADIALGEALLASOKNRIEHQYV 246
    || ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 GAFAPICNTPERLFQRTQLGVCSSEALATPRRAASSARMEIERDLTTSKRDLESIIVR 349
OY 247 ASITTRLDVYTSLSKYPAMPSSLKKNQOVHLYPTITGDIAHLVTAIVDRLHPTPALGG 306
    : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 ENIREKINGICDRVVVYVQKTVKRLARVQHLVLSQLAGRLRKEDDEXKILALHPTPAVCG 409
OY 307 VPREALIYIATHEKTPRGIFAGPIGFTADNGSEFVVGIRSMYVNOTORRALLFAGAGI 366
    : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 LPAEEALLIKETIESFDRCMGVAGPIGFFGGEES-EFVAGIRSAIV-EKGIGALIVAGTGI 467
OY 367 VADSDAQOYEYETGLK 382
    || || : : : : :
Db 468 VAGSDPSSEMMELDLK 483

RESULT 7
ID 09S7H8 PRELIMINARY: PRT: 503 AA.
AC 09S7H8:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE ISOCHORISMATE SYNTHASE (ICSI).
GN FIM20.39 OR F25A4.31.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC Brassicaceae; Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA:
RA Llin X., Keul S., Koon C.D., Benito M., Creasy T.H., Haas B.,
RA Rinning C.M., Koo H., Fujil C.Y., Uterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome 1 BAC FIM20 genomic sequence.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA:
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Lee J.M., Li J., Gonzalez A., Liu A., Liu K., Vaysberg M., Sakano H.,
RA Chin C., Choi E., Chou J., Altafi H., Araujo R., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Hong B., Huizer L., Khan S., Kim C., Palm C., Rowley D., Shinn P.,
RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases."
RL EMBL: AC011765; AAF15941.1; -.
DR EMBL: AC008263; AAD55272.1; -.
DR INTERPRO: IPR000350; -.
DR PIRAM: PF00425; chorismate_bind.1.
DR PRODOM: PD000779; -.
SQ SEQUENCE 503 AA: 55268 MW: 8CFC36477D54F61E CRC64;

Query Match 15.3%; Score 307.5; DB 10; Length 503;
Best Local Similarity 28.2%; Pred. No. 8,6e-16;
Matches 106; Conservative 59; Mismatches 158; Indels 53; Gaps 11;

OY 21 ALETFEFGA-YFATPADTLRFGIGAIAATAKTA-----QALQGAVFAGQSFDEQYVPS 73
    : : ||| ||| : : : : : : : : : : : : : : : : : : : : :
Db 147 AVEWMPFGAFYGSVPOVENREGGSSMLAATIAMDELSTLEMALEAQ----- 196
OY 74 ELMAGFVWPEVMTIADKITFGSDTVVSDFTTWLAQVVPKOPNTVYTSHTDEVDMIR 133
    : : : : : : : : : : : : : : : : : : : : : : : : :
Db 197 -----ETMLQVSSVVMKLRNRSIG-----VSLSKNHVPKGAIFYPA 233
OY 134 TENLIDTL-AIDQTLAKYVFGQOTLOLSDT-----LRLAQITRLALAEANHYHVLK 187
    : : : : : ||| : : : : : : : : : : : : : : : : : :
Db 234 VEKALEMINOKSSPLNKVVLARNRI-TTDTIDPIAMLAQLOR---EGHDVAFCLOPP 289
```

```
OY 188 D-ELFISATPERLVAMSGGQIATAVAGTSRRGTGADGADIALGEALLASOKNRIEHQYV 246
    || ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 GAFAPICNTPERLFQRTQLGVCSSEALATPRRAASSARMEIERDLTTSKRDLESIIVR 349
OY 247 ASITTRLDVYTSLSKYPAMPSSLKKNQOVHLYPTITGDIAHLVTAIVDRLHPTPALGG 306
    : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 ENIREKINGICDRVVVYVQKTVKRLARVQHLVLSQLAGRLRKEDDEXKILALHPTPAVCG 409
OY 307 VPREALIYIATHEKTPRGIFAGPIGFTADNGSEFVVGIRSMYVNOTORRALLFAGAGI 366
    : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 LPAEEALLIKETIESFDRCMGVAGPIGFFGGEES-EFVAGIRSAIV-EKGIGALIVAGTGI 467
OY 367 VADSDAQOYEYETGLK 382
    || || : : : : :
Db 468 VAGSDPSSEMMELDLK 483

RESULT 8
ID P95475 PRELIMINARY: PRT: 391 AA.
AC P95475:
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE ISOCHORISMATE SYNTHASE.
GN FBSC.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WCS374;
RA Mercado-Blanco J., Olsson P.E., van Loon L.C., Bakker P.A.H.M.;
RT "Analysis of a gene cluster involved in salicylate biosynthesis and
RT detection of a novel siderophore in Pseudomonas fluorescens WCS374.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y09356; CAA70528.1; -.
DR INTERPRO: IPR000350; -.
DR PIRAM: PF00425; chorismate_bind.1.
DR PRODOM: PD000779; -.
SQ SEQUENCE 391 AA: 43102 MW: 6654B0901C7D56A1 CRC64;

Query Match 15.2%; Score 304.5; DB 2; Length 391;
Best Local Similarity 28.6%; Pred. No. 1e-15;
Matches 108; Conservative 46; Mismatches 160; Indels 63; Gaps 10;

OY 68 QEYPOSELAGFVFP---EVMTIADKITE-----GSDTVSDFTTWLAQ----- 110
    : : : : : : : : : : : : : : : : : : : : : : : : :
Db 12 EEVQLAEVQRSFSTGSDRELAVTGMLORETPALIGDDANSLFQQTINQALDRAREAG 71
OY 111 -----FVVKOPNTVYTSHTDEVDMIER-----TENLIDTLAID 144
    : : : : : : : : : : : : : : : : : : : : : : : : :
Db 72 SNEIIVGAIPIFPAPERSCL---YIPEHAQWRTRDIHAKTGMSTLPRLIEQKNIPDEQARK 128
OY 145 QTLAKYVFG-----RQOTLOLSDTLRLAQIIRALDAQNT-----YHVLKRRD-EL 190
Db 129 RAVEHAIVVFRISDVAKAVSVQRELIFANDVDVSAQLQHLKAKQNSGTHFRVPMPDGTT 188
OY 191 FTSATPERLVAMSGGQIATAVAGTSRRGTGADGADIALGEALLASOKNRIEHQYVAST 250
    : : ||| ||| : : : : : : : : : : : : : : : : : : : : :
Db 189 LGVSPPELLVRRREGSLSSNPLAGSKAKRNADPADRRNMDWLTSKDIYEHQFYQDIV 248
OY 251 TRLDVYTSLSKYPAMPSSLKKNQOVHLYPTITGDIA-AHLSTVAVDRLHPTPALGVPR 309
    : : : : : : : : : : : : : : : : : : : : : : : : :
Db 249 SOLGKICTOLNVPORSLSTPALMHLSTRIGETLADPAVSALQLCRLHPTPAVCGPPT 308
OY 310 EALYIATHEKTPRGIFAGPIGFTADNGSEFVVGIRSMYVNOTORRALLFAGAGI 369
    : : ||| ||| : : : : : : : : : : : : : : : : : : : : :
Db 309 EAKRRLIRFVEPERGLFGLVGMCDAGNGEWMYVTRKGVRRNKR--LFGAGIVFA 366
OY 370 SDAQOYEYETGLKFEPM 386
```



Best Local Similarity 26.3%; Pred. No. 1.4e-09;  
Matches 74; Conservative 50; Mismatches 141; Indels 16; Gaps 5;

QY 108 LAQVPRKPNVTSHYTDV-----DWIERTENLIDTLAIDDTLAKVYGRROOTLDS 161  
DB LTRVAVAPPALPSELPSEYALMGDFODAVEDIKERIAGAEFOVYPSQRETPCTA 267  
QY 162 DTLRLAIIIRALAEQANTYHVHLKRHDELFTSATPERLVAMSGOIAATAVAGSRGCTD 221  
DB 268 SALDQVYVLRATNSPYMYLLRLDGF--VYGSSEALYKVEDGRAMVHPIAGTRPRCAT 325  
QY 222 GADDIALGEALLASOKNRIEHQYVASTITRLQDVT--SLKVPAMPSLKNKOYOLYT 279  
DB 326 PREQDSIADELADPKERAHEMLMLVDLGRNDLGRVCEGSEVYEDFMSVERYSHVMIHS 385  
QY 280 PINGDIAHLSVTAIVDRLHPTALGVPREALITYATHTKTRGLGAGPIGYFTADNS 339  
DB 386 TVIGRVAHPGRAPDVLRACFPAGTLGAPKPRALQIIDELEPSRGLYGGCVGLDFAGD 445  
QY 340 GEFVVGIRSMVYNOTORATLF--AGAGIVADSDAOEYEE 378  
DB 446 SDFIAIRFTALL----RDGTAYVQAGACVADSDPAEDTE 482

## RESULT 12

ID P72539 PRELIMINARY; PRT; 719 AA.  
AC P72539;  
DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE PAPA.  
GN PAPA.  
OS Streptomyces pristinaespiralis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 97197164.  
RA Blanc V., Gall P., Bamas-Jacques N., Lorenzon S., Zagorec M.,  
RA Schluniger J., Bisch D., Blanche F., Debussche L., Crouzet J.,  
RA Thibaut D.  
RT "Identification and analysis of genes from Streptomyces  
RT pristinaespiralis encoding enzymes involved in the biosynthesis of the  
RT 4-dimethylamino-L-phenylalanine precursor of pristinamycin I.";  
RL MOL. Microbiol. 23:191-202(1997).  
DR EMBL: U60417; AAC44866.1; -.  
DR INTERPRO: IPR000350; -.  
DR INTERPRO: IPR000991; -.  
DR INTERPRO: IPR002385; -.  
DR PFAM: PF00425; chorismate\_bind; 1.  
DR PFAM: PF00117; GATase; 1.  
DR PRINTS: PRO0095; ANTSMTHASEI.  
DR PRINTS: PRO0096; GATASE.  
DR PRINTS: PRO0097; ANTSMTHASEII.  
DR PROSITE: PS00442; GATASE\_TYPE\_1; 1.  
DR PRODOM: PD00779; -; 1.  
SQ SEQUENCE 719 AA; 77363 MW; BFEBCBDA91D8368 CRC64;

Query Match 10.9%; Score 219; DB 2; Length 719;  
Best Local Similarity 25.7%; Pred. No. 1e-08;  
Matches 77; Conservative 53; Mismatches 142; Indels 28; Gaps 7;

QY 101 VSDPTTLAQFVPRKQPMVTSHYTDV-----WIERTENLIT--DTLAIQD 145  
DB 402 LTTDAARTLATTAAPRPTTLPPDQLPALDVYRHSLSRPRYRELVEECRLITDGETYEVL 461  
QY 146 TLAKVVFGRROOTLQSDTLRLAIIIRALAEQANTYHVHLKRHDELFTSATPERLVAM-SG 204  
DB 462 TNNLRVGR-----IDPLRAYRALTIVSPA--PYAAYLDPPGATVLSSEERLRLGAD 513  
QY 205 GOIATAVAGTSRGTGDADDIALGEALLASOKNRIEHQYVASTITRLQDVT--SLKV 262

DB 514 GWAEKPIKGTRRPRCAGPAQAOAAVAVSLAAAEKRESEMLIYDLVRNDLGVCDIGSVHV 573  
QY 263 PAMPSLKNKOYOLHTYITGDIAHLSVTAIVDRLHPTALGVPREALITYATHTK 322  
DB 574 PGLFEVETTYATVHOLVSTRGRLAADVSRPRAVAFGSGMTGAPKVRTQOFIDRLK 633

QY 323 PRGFAGPIGYFTADNSGEFVVGIRSMVYNOTORATLFPAGAGIYADSDAOEYEEGLK 382  
DB 634 PRGVYSGALGIFALSGAADLSIVRTIYA--TEEAATIGVGAVVALSDPDDEVEREMLLK 691

## RESULT 13

ID 094582 PRELIMINARY; PRT; 489 AA.  
AC 094582;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE ANTHRANILATE SYNTHASE COMPONENT I.  
GN SPC1442.09  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;  
OC Schizosaccharomycetaceae; Schizosaccharomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972H-;  
RA Lyne M., Rajandream M.A., Barrell B.G., Voiclaert G.,  
RL Submitted (Oct-1998) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AL031966; CA421443.1; -.  
DR INTERPRO: IPR000350; -.  
DR PFAM: PF00425; chorismate\_bind; 1.  
DR PRINTS: PRO0095; ANTSMTHASEI.  
DR PRODOM: PD00779; -; 1.  
SQ SEQUENCE 489 AA; 54959 MW; 442D522BA59EBA31 CRC64;

Query Match 10.9%; Score 218.5; DB 3; Length 489;  
Best Local Similarity 33.7%; Pred. No. 6.2e-09;  
Matches 65; Conservative 35; Mismatches 84; Indels 9; Gaps 5;

QY 192 ISAPPERLVAMSGOIAATAVAGTSRGTGDADDIALGEALLASOKNRIEHQYV--AS 248  
DB 275 IGASPELLVKESEHRIINHPIAGTVPRCKTKEDEAYAKDLLASVYKDBAEHVMLVDLARN 334  
QY 249 ITTRLDQVTSLSKVPAMPSLKNKOYOLHTYITGDIAHLSVTAIVDRLHPTALGVP 308  
DB 335 DVSRYCDLDT--SYDKLMTIEKFSHVQHLVSQVSVLPRDKTRFDARSRIFPAGTVSGSP 393  
QY 309 REAALYVATHEKTPRGLFAGPIG-YFPADNSGEFVVGIRSM-VYNOTORATLFPAGAGI 366  
DB 394 KYRALQLYVGLKEKERGIYACAGVGRMGYEDDNMDICAIRIMVYKDG--VYLDAGGCI 450  
QY 367 VADSDAOEYEE 379  
DB 451 VFSDQDEQVEYET 463

## RESULT 14

ID 09X918 PRELIMINARY; PRT; 434 AA.  
AC 09X918;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE SALICYLATE SYNTHETASE, IRP9.  
GN IRP9.  
OS *Shigella enterocolitica*.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC *Yersinia*.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rakin A.V., Pelludat C., Noelling C., Schubert S., Jacobi C.,



---

**THIS PAGE BLANK (USPTO)**